

Artikel

by KKMMS 1

Submission date: 17-Jan-2023 09:35PM (UTC-0600)

Submission ID: 1994541360

File name: 57300-75676677337-1-PB.pdf (471.56K)

Word count: 5964

Character count: 37539



Perbandingan *Particle Swarm Optimization* dengan *Genetic Algorithm* dalam *Feature Selection* untuk Analisis Sentimen pada Permendikbudristek PPKS-LPT

Gregorius Guntur Sunardi Putra^{#1}, Windra Swastika^{#2}, Paulus Lucky Tirma Irawan^{#3}

[#]Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Ma Chung
Villa Puncak Tidar Blok N no. 1, Malang 65151

¹311810015@student.machung.ac.id

²windra.swastika@machung.ac.id

³paulus.lucky@machung.ac.id

Abstrak— Kekerasan seksual merupakan salah satu masalah serius yang perlu diselesaikan. Setiap tahunnya terdapat lebih dari 2000 aduan kekerasan seksual. Di satuan Pendidikan, sebanyak 55% dari pelaku merupakan guru atau tenaga pendidik. Jumlah yang ada masih belum merepresentasikan jumlah sebenarnya karena 63% dari peserta survei DIKTI tidak melaporkan kasus kekerasan seksual yang diketahuinya dengan alasan menjaga nama baik kampus. Dengan begitu, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemendikbud) menerbitkan Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi (Permendikbudristek) Nomor 30 tahun 2021 tentang Pencegahan dan Penanganan Kekerasan Seksual di Lingkungan Perguruan Tinggi (PPKS-LPT). Peraturan ini menimbulkan kontroversi di masyarakat sehingga perlu dilakukan klasifikasi sentimen. Klasifikasi sentimen dilakukan menggunakan algoritma *Multinomial Naïve Bayes*, *Complement Naïve Bayes*, *Support Vector Machine*, dan *Nearest Neighbor* dengan menggunakan data seleksi fitur *Genetic Algorithm* dan *Particle Swarm Optimization* pada data hasil ekstraksi fitur TF-IDF dari data Twitter yang diambil pada tanggal 1 September-30 Desember 2021 menggunakan kata kunci “permendikbud”, hashtag “cabutpermendikbudristekno30”, dan hashtag “dukungpermendikbud30”. Kemudian klasifikasi akan dievaluasi menggunakan *10-Fold Cross Validation*. Hasil dari penelitian ini, algoritma CNB memiliki kinerja paling baik jika dibandingkan dengan 4 algoritma klasifikasi lain dengan nilai *accuracy* 0.688, *precision* 0.685, *recall* 0.688, dan *f1-score* 0.683. Seleksi fitur GA belum mampu menemukan fitur-fitur yang memiliki akurasi lebih baik jika dibandingkan menggunakan semua fitur, sedangkan seleksi fitur PSO mampu menemukan fitur yang memiliki akurasi sama atau lebih baik jika dibandingkan menggunakan semua fitur dengan peningkatan tertinggi sebanyak 0.8% pada MNB. Sebanyak 49% dari responden merupakan pendukung pengesahan Permendikbudristek PPKS-LPT.

Kata kunci— *Feature Selection*, *Genetic Algorithm*, *Machine Learning*, *Particle Swarm Optimization*, *Permendikbudristek PPKS*, *Sentiment Analysis*.

I. PENDAHULUAN

Kekerasan seksual merupakan salah satu masalah serius yang perlu diselesaikan. Kekerasan seksual terdiri dari 13 jenis kekerasan seperti pelecehan dan pemerkosaan [1]. Komnas Perempuan mendapatkan lebih dari 2000 aduan mengenai kasus kekerasan seksual dan terus meningkat dalam 5 tahun terakhir [2]. Dari semua aduan, mayoritas pelaku merupakan orang terdekat dari korban seperti keluarga, pacar, teman dan tenaga pendidik. Tahun 2021, sebanyak 55% pelaku kekerasan seksual di satuan pendidikan merupakan guru [3]. Data yang dilaporkan bukan merupakan keseluruhan dari kekerasan seksual yang terjadi dikarenakan 63% kasus kekerasan seksual tidak dilaporkan untuk menjaga nama baik kampus [4]. Oleh karena itu, data kekerasan seksual masih belum mencerminkan jumlah kasus sebenarnya dan mungkin lebih banyak dari yang dilaporkan.

Secara khusus kekerasan seksual tidak memiliki ketetapan hukum yang pasti, sehingga ketika terjadi kekerasan seksual maka akan mengacu pada KUHP dan UU Perkawinan, dimana ruang lingkup kekerasan masih terbatas pada perzinaan dan perselingkuhan. Untuk mengisi kekosongan hukum, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi menerbitkan Permendikbudristek No. 30 tahun 2021 tentang Pencegahan dan Penanganan Kekerasan Seksual di Lingkungan Perguruan Tinggi pada bulan September 2021. Peraturan ini secara khusus mengatur tentang kekerasan seksual di lingkungan perguruan tinggi dengan harapan sebagai acuan hukum jika terjadinya kekerasan seksual di kampus [5]. Beberapa bulan setelah disahkan, Permendikbudristek memicu kritikan dan menjadi

perbincangan oleh berbagai kalangan, terutama di media sosial.

Salah satu media sosial yang sering digunakan untuk beropini, berpendapat, dan berkomentar tentang suatu peristiwa yang terjadi adalah Twitter. Dengan batasan 280 karakter, pengguna dapat mencurahkan segala pemikiran atau pandangan terhadap suatu fenomena yang terjadi di masyarakat. Media sosial Twitter memiliki pengguna aktif sebanyak 173 juta pada tahun 2021, sehingga menjadi representasi dari pandangan masyarakat Indonesia secara umum [6]. Jumlah pengguna setara dengan 63,6% dari populasi penduduk di Indonesia, sehingga dapat dimanfaatkan untuk mengetahui sentimen masyarakat mengenai Permendikbudristek PPKS-LPT.

Setiap individu mempunyai padangan, pendapat, dan opini yang berbeda-beda terhadap suatu peristiwa dan diungkapkan dengan cara yang berbeda pula. Saat ini, individu sering kali mengungkapkan pandangan ke dalam media sosial dalam bentuk teks. Metode yang dilakukan untuk menganalisis sentimen dari individu tersebut berdasarkan teks yang ditulis adalah metode *Sentiment Analysis*. Metode ini melakukan analisis pada data teks berdasarkan kamus atau menggunakan pembelajaran mesin [7]. Pembagian sentimen ke dalam kelas positif, negatif, dan netral, dapat dilakukan secara manual, namun dengan jumlah data yang besar, hal ini membutuhkan usaha yang lebih besar. Oleh karena itu, perlu adanya sebuah mesin yang dapat mengklasifikasi *tweet* ke dalam sentimen positif, negatif, dan netral secara otomatis.

Pengklasifikasian secara otomatis dapat dilakukan dengan menggunakan *machine learning*. Penelitian [8] menggunakan SVM untuk menilai sentimen kuota Mendikbud dalam Permendikbud Nomor 19 Tahun 2020. Labelisasi yang dilakukan menggunakan metode *lexicon based* yang memisahkan data positif, negatif dan netral berdasarkan koleksi korpus positif dan negatif. Data diambil dari komentar di akun Instagram @kemendikbud pada tanggal 27 Agustus – 30 September 2020 menghasilkan 4520 komentar. Hasil Penelitian ini menunjukkan 70,67% bersentimen positif dengan kinerja akurasi SVM RBF 79,67%. Studi oleh [9] menggunakan *BernoulliNB*, *Logistic Regression*, dan SVM untuk mengelompokkan data sentimen RUU penghapusan kekerasan seksual ke dalam kelas positif, negatif, dan netral. Hasil dari studi ini menyatakan bahwa *BernoulliNB* menjadi model paling baik dengan kinerja *f1-score* 0,65. Penelitian yang dilakukan [10] menggunakan NB, SVM, KNN, dan ANN menghasilkan akurasi terbaik pada SVM dengan nilai 94%. Sedangkan pada [11] menggunakan KNN untuk melakukan analisis sentimen mampu menghasilkan akurasi sebesar 87% dan pada [12] mampu menghasilkan akurasi sebesar 67,2%. Studi yang serupa dilakukan oleh [13] menggunakan KNN dengan pembobotan TF-IDF mampu menghasilkan akurasi sebesar 84,65%.

Jumlah fitur yang besar akan memperlambat algoritma klasifikasi dan mengurangi kinerja yang disebabkan tidak optimalnya sebuah algoritma klasifikasi. Pengoptimalan

fitur dinamakan dengan *feature selection*, dimana metode ini berusaha menggunakan fitur yang memiliki pengaruh pada kinerja algoritma klasifikasi [14]. Penelitian [15] menggunakan seleksi fitur *Genetic Algorithm* (GA) pada algoritma *Support Vector Machine* (SVM) menghasilkan peningkatan akurasi sebesar 15,76% dibandingkan dengan menggunakan semua fitur. Penelitian serupa dilakukan oleh [16] yang menggunakan GA untuk seleksi fitur pada SVM, *AdaBoost*, dan KNN menghasilkan peningkatan akurasi tertinggi pada SVM sebanyak 0,77%. Studi [17] membandingkan seleksi fitur GA pada KNN, NC, ANN, RF, dan *Rpart* menghasilkan peningkatan tertinggi pada RF sebesar 0,92%. Penelitian oleh [18] menghasilkan seleksi fitur PSO mampu meningkatkan klasifikasi *Naive Bayes* sebanyak 4%.

Dalam studi ini, menggunakan algoritma metaheuristik untuk meningkatkan hasil akurasi pada algoritma klasifikasi. Algoritma metaheuristik yang digunakan adalah algoritma *Particle Swarm Optimization* dan *Genetic Algorithm*. Algoritma klasifikasi yang dilakukan percobaan yaitu *Multinomial Naive Bayes*, *Complement Naive Bayes*, *Support Vector Machine*, dan *Nearest Neighbor*. Setiap percobaan akan dievaluasi berdasarkan nilai *accuracy*, *precision*, *recall*, dan *f1-score*.

II. LANDASAN TEORI

A. TF-IDF

Term Frequency – Inverse Document Frequency (TF-IDF) merupakan metode pembobotan kata dengan mengekstraksi fitur dari suatu teks dengan mempertimbangkan kemunculan kata pada sebuah dokumen dan ukuran kepentingan kata tersebut [9]. Perhitungan rumus TF-IDF dapat dihitung menggunakan persamaan berikut:

$$TF = \frac{\text{frekuensi term pada tweet}}{\text{total term pada tweet}} \quad (1)$$

$$IDF = \log \frac{\text{total tweet}}{\text{frekuensi tweet yang mengandung term}} \quad (2)$$

$$TF-IDF = TF \times IDF \quad (3)$$

B. Seleksi Fitur

Seleksi fitur merupakan proses yang memilih fitur-fitur penting yang akan digunakan dalam klasifikasi dengan tujuan meningkatkan akurasi [14]. Melakukan seleksi fitur pada data set dapat mengurangi *runtime* secara signifikan [19].

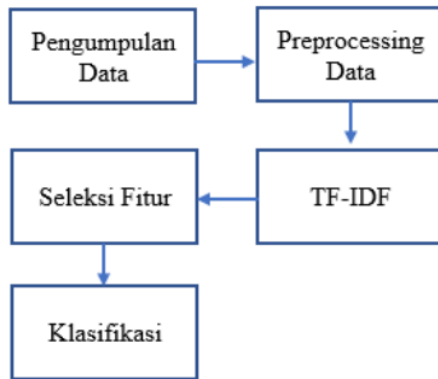
C. Klasifikasi

Klasifikasi merupakan sebuah analisis data yang mengekstrak suatu model untuk merepresentasikan kelas data. Model yang disusun mencakup pengklasifikasian serta prediksi kategori kelas [20].

III. METODE PENELITIAN

Dalam melakukan penelitian ini, terdapat tahapan penelitian seperti yang terlihat pada Gambar 1. Tahapan

dimulai dari pengumpulan data Twitter menggunakan API Twitter, *preprocessing* data untuk menyesuaikan standar, pembobotan data dengan TF-IDF, seleksi fitur pada data, dan pembuatan model klasifikasi.



Gambar 1 Tahapan penelitian

A. Pengumpulan Data

Pengumpulan data dilakukan dengan mengambil tweet berbahasa Indonesia menggunakan API Twitter. Pengambilan menggunakan kata kunci permendikbud, dukungpermendikbud, dan cabutpermendikbud yang diambil dalam rentang waktu 1 September 2021 sampai 30 Desember 2021. Hasil yang didapatkan sebanyak 20.461 *tweet*. Pada Tabel I merupakan sampel 3 data dari seluruh data yang didapatkan.

TABEL I
HASIL DATA YANG DIKUMPULKAN

| Date | Tweet | User |
|-------------|--|---------------|
| 25 Oct 2021 | Sehingga mahasiswa mendesak agar segera disahkan Permendikbud Ristek PPKS demi korban KS #univeristasbrawijaya #kekerasanseksual #pemberdayaanperempuanprogesif #permendikbudristekPPKS #emub2021 | EM_UBOfficial |
| 26 Oct 2021 | Ini merupakan langkah besar untuk menciptakan lingkungan akademik yang ramah gender Materi dan info lebih lanjut tentang Permen PPKS dapat diakses di link ini bit.ly/PermenPPKS | airakusumas |
| 03 Nov 2021 | Ikadi Bersama 12 Ormas Tolak Permendikbud No 30 2021 Tentang Kekerasan Seksual Al Karena Dinilai Bertentangan Dg Pancasila UUDNRI 1945 psl 31 ayat 3 Mendikbud Agar Segera Mencabut amp Menggantinya Dg Yg Sesuai Pancasila amp UUDNRI 1945 dan Budaya Indonesia | hnurwahid |

Dari hasil yang didapatkan kemudian dipilih berdasarkan kedekatan hubungan pembahasan dengan Permendikbud PPKS dan menggunakan Bahasa Indonesia. Data yang dihapus seperti penjualan barang, kata-kata bijak, *tweet* yang duplikat, dan kalimat yang ditulis dalam Bahasa Inggris. Hasil setelah penghapusan data menyisakan 15.637 data *tweet*.

Setelah data yang sudah dikurangi kemudian diberikan label secara manual untuk menentukan sentimen tersebut positif, negatif, dan netral. Pemberian label dilakukan berdasarkan nuansa dari sentimen terhadap permendikbud. Pada Tabel II merupakan contoh dari sentimen positif, negatif, dan netral pada data set.

TABEL II
PEMBERIAN LABEL PADA DATA TWEET

| No | Tweet | Sentimen |
|----|--|----------|
| 1 | Sehingga mahasiswa mendesak agar segera disahkan Permendikbud Ristek PPKS demi korban KS #univeristasbrawijaya #kekerasanseksual #permendikbudristekPPKS | Positif |
| 2 | Ini merupakan langkah besar untuk menciptakan lingkungan akademik yang ramah gender Materi dan info lebih lanjut tentang Permen PPKS dapat diakses di link ini bit.ly/PermenPPKS | Netral |
| 3 | Ikadi Bersama 12 Ormas Tolak Permendikbud No 30 2021 Tentang Kekerasan Seksual Al Karena Dinilai Bertentangan Dg Pancasila UUDNRI 1945 psl 31 ayat 3 Mendikbud Agar Segera Mencabut amp Menggantinya Dg Yg Sesuai Pancasila amp UUDNRI 1945 dan Budaya Indonesia | Negatif |

B. Preprocessing Data

Tahap *Preprocessing* data merupakan tahapan yang berfungsi untuk mengubah hasil data mentah yang tidak terstruktur menjadi data yang terstruktur. Dalam data teks, *preprocessing* memiliki beberapa tahap yaitu *case folding*, *cleaning*, tokenisasi, normalisasi, *stopword*, *stemming*, pembobotan kata, dan seleksi fitur. Data hasil *preprocessing* siap digunakan untuk diklasifikasi menggunakan *machine learning*.

1) *Case Folding*: Data yang sudah dikumpulkan kemudian dilakukan *case folding* yaitu tahap dimana semua huruf kapital pada teks menjadi huruf kecil. Tabel III merupakan perbandingan sebelum dan sesudah dilakukan tahap *case folding*.

TABEL III
PERBANDINGAN SEBELUM DAN SETELAH TAHAP CASE FOLDING

| Sebelum | Sesudah |
|--|--|
| Sehingga mahasiswa mendesak agar segera disahkan Permendikbud Ristek PPKS demi korban KS #univeristasbrawijaya #kekerasanseksual | sehingga mahasiswa mendesak agar segera disahkan permendikbud ristek ppsk demi korban ks #univeristasbrawijaya #kekerasanseksual |

| Sebelum | Sesudah |
|-------------------------|-------------------------|
| #permendikbudristekPPKS | #permendikbudristekppks |

2) *Cleaning*: Data yang sudah dikumpulkan kemudian dilakukan penghapusan karakter simbol, *mention* ('@') nama pengguna, *hashtag* ('#'), karakter numerik, dan *link url*. Hasil *cleaning* dilihat pada Tabel IV.

TABEL IV
PERBANDINGAN SEBELUM DAN SETELAH TAHAP *CLEANING*

| Sebelum | Sesudah |
|--|--|
| sehingga mahasiswa mendesak agar segera disahkan perpendikbud ristek ppks demi korban ks #univeristasbrawijaya #kekerasanseksual #permendikbudristekppks | sehingga mahasiswa mendesak agar segera disahkan perpendikbud ristek ppks demi korban ks |

3) *Tokenisasi*: Melakukan pemotongan kata menjadi *token* pada setiap *tweet* yang menghasilkan satu potongan kata tunggal. Hasil *Tokenisasi* dapat dilihat pada Tabel V.

TABEL V
PERBANDINGAN SEBELUM DAN SETELAH TAHAP *TOKENISASI*

| Sebelum | Sesudah |
|--|--|
| sehingga mahasiswa mendesak agar segera disahkan perpendikbud ristek ppks demi korban ks | sehingga mahasiswa mendesak agar segera disahkan perpendikbud ristek ppks demi korban ks |

4) *Normalisasi*: Memperbaiki kata-kata yang salah dalam teks, dan menyamakan perbedaan penulisan untuk arti kata yang sama berdasarkan korpus yang dibuat. Hasil *Normalisasi* dapat dilihat pada Tabel VI.

TABEL VI
PERBANDINGAN SEBELUM DAN SETELAH TAHAP *NORMALISASI*

| Sebelum | Sesudah |
|--|---|
| sehingga mahasiswa mendesak agar segera disahkan perpendikbud ristek ppks demi korban ks | sehingga mahasiswa mendesak agar segera disahkan perpendikbud ristek ppks demi korban kekerasan seksual |

5) *Stopword*: Menghapus kata-kata yang tidak penting yang memiliki kemunculan paling banyak seperti notasi HTML, kata hubung, kata imbuhan, kata ganti subjek, dan kata-kata yang tidak penting lainnya yang tidak merepresentasikan apa pun secara jelas. Korpus *Stopword* menggunakan *stopword* pada Sklearn dan Sastrawi. Hasil *Stopword* dapat dilihat Pada Tabel VII.

TABEL VII
PERBANDINGAN SEBELUM DAN SETELAH TAHAP *STOPWORD*

| Sebelum | Sesudah |
|---|---|
| sehingga mahasiswa mendesak agar segera disahkan perpendikbud ristek ppks demi korban kekerasan seksual | mahasiswa mendesak disahkan perpendikbud ristek ppks korban kekerasan seksual |

6) *Stemming*: Hasil yang telah dilakukan *stopword* kemudian dilakukan proses *stemming* yaitu pemotongan dokumen dengan menghilangkan kata imbuhan, sisipan, awalan, dan akhiran untuk menghasilkan kata dasar. Pengubahan sebuah kata tunggal menjadi kata dasar ini dibuat dengan berdasarkan Bahasa Indonesia yang baik dan benar. Hasil *Stemming* dapat dilihat pada Tabel VIII.

TABEL VIII
PERBANDINGAN SEBELUM DAN SETELAH TAHAP *STEMMING*

| Sebelum | Sesudah |
|---|---|
| mahasiswa mendesak disahkan perpendikbud ristek ppks korban kekerasan seksual | mahasiswa desak sah perpendikbud ristek ppks korban keras seksual |

C. *TF-IDF*

Hasil yang sudah dilakukan oleh *preprocessing*, kemudian dilakukan proses pembobotan *Term Frequency – Inverse Document Frequency (TF-IDF)*. Perhitungan bobot ini menggunakan *vector* yang bernilai dalam rentang 0 sampai 1 dengan mempertimbangkan frekuensi kemunculan kata [9]. Penelitian ini menggunakan *library sklearn* untuk menghitung nilai *TF-IDF* dalam bentuk *vector*.

D. *Seleksi Fitur*

Seleksi fitur menggunakan algoritma *metaheuristik* sebuah algoritma yang mengambil konsep dari alam. Setiap fitur akan direpresentasikan ke dalam nilai 0 untuk

tidak menggunakan fitur dan 1 untuk menggunakan fitur, Berikut metode seleksi fitur yang digunakan:

1) *Particle Swarm Optimization*: PSO merupakan salah satu algoritma yang dapat digunakan untuk seleksi fitur yang secara otomatis dapat mengembangkan jumlah fitur digunakan lebih sedikit dengan akurasi yang lebih besar dibandingkan dengan menggunakan semua fitur [15]. Posisi setiap partikel akan berada dalam rentang nilai 0 dan 1 sehingga setiap posisi akan dilakukan *threshold* menggunakan persamaan (4).

$$threshold = x \begin{cases} 1, & \text{if } xi > 0.5 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Tahapan PSO dalam melakukan seleksi fitur pada data set yang telah diekstraksi fitur sebagai berikut:

- Menentukan populasi nilai posisi setiap partikel secara acak dalam rentang nilai 0 sampai 1.
- Menghitung nilai *threshold* dari setiap posisi partikel dan nilai 1 sebagai fitur terpilih pada data ekstraksi fitur.
- Menghitung nilai *fitness* setiap partikel menggunakan nilai evaluasi dari algoritma klasifikasi menggunakan data fitur terpilih pada data ekstraksi fitur.
- Nilai *fitness* terbaik dalam satu populasi akan menjadi *Global Best*. Jika mencapai kondisi yang ditentukan maka tahapan berhenti langkah d, jika belum lanjutkan ke langkah e.
- Mengubah nilai kecepatan dan posisi setiap partikel dan mengulangi langkah b sampai mencapai kondisi yang ditetapkan.

2) *Genetic Algorithm*: Algoritma genetika sering digunakan untuk menyelesaikan masalah optimasi yang nilainya belum jelas[16]. Tujuan utama GA adalah untuk menemukan solusi optimal dalam serangkaian solusi potensial. Algoritma ini berpotensi dalam menghasilkan pemilihan fitur yang lebih baik dan sekaligus menjadi parameter yang optimal. Tahapan algoritma genetika dalam seleksi fitur sebagai berikut:

- Menentukan populasi kromosom awal secara acak dengan nilai gen dalam bentuk biner 0 dan 1.
- Menghitung nilai *fitness* setiap kromosom dengan menggunakan evaluasi dari algoritma klasifikasi pada data ekstraksi fitur menggunakan fitur terpilih dari setiap kromosom.
- Memilih kromosom sebagai induk, dalam penelitian ini menggunakan *tournament selection* dengan nilai K sebanyak 3.
- Melakukan kombinasi pada dua induk kromosom dengan titik pemisahan secara acak untuk mendapatkan dua kromosom baru. Dalam penelitian ini menggunakan metode kombinasi *one-point*.
- Mutasi gen pada induk kromosom secara acak untuk mendapatkan kromosom baru. Dalam penelitian ini menggunakan mutasi bit-string yang mengubah nilai gen 0 menjadi gen 1 dan sebaliknya.
- Menghitung nilai *fitness* setiap populasi kromosom baru yang didapatkan dari proses kombinasi dan

mutasi. Nilai *fitness* dihitung dengan cara yang sama seperti langkah b. Kromosom yang memiliki nilai *fitness* lebih baik akan digunakan dalam generasi selanjutnya.

- Langkah c sampai g akan diulangi sampai mencapai kondisi tertentu yang ditetapkan.

E. Klasifikasi

Ada beberapa algoritma klasifikasi yang akan digunakan yaitu:

1) *Multinomial Naïve Bayes*: merupakan sebuah algoritma yang berdasarkan pada *Teorema Bayes*. Algoritma ini melakukan prediksi peluang terjadinya suatu peristiwa berdasarkan peristiwa yang pernah terjadi menggunakan persamaan (5). Metode ini mampu menghasilkan kinerja yang bagus pada jumlah data latih yang sedikit. Pembangunan model MNB menggunakan *library Sk-Learn* pada fungsi *MultinomialNB*.

$$MNB = \mathop{\text{argmax}}_{\text{label docd}} \frac{\text{ldocs } j!}{\text{all docd}} * \prod \frac{n_k+1}{n+|\text{kosakata}|} \quad (5)$$

2) *Complement Naïve Bayes*: dikembangkan dari MNB yang memiliki kelemahan dalam menangani data tidak seimbang, sehingga CNB dibuat untuk mengoptimalkan kinerja dari *teorema bayes* pada data teks ketika data pelatihan yang digunakan tidak seimbang. Algoritma ini melakukan inversi pada perhitungan MNB seperti pada persamaan (6).

$$CNB = \mathop{\text{argmax}}_{\text{label docd}} \frac{\text{ldocs } j!}{\text{all docd}} * \prod \frac{n+|\text{kosakata}|}{n_k+1} \quad (6)$$

3) *Support Vector Machine*: merupakan metode klasifikasi yang berusaha mencari *hyperplane* yang memisahkan dua data berbeda. Model awal SVM ini, hanya untuk 2 kelas berbeda, dikarenakan penelitian ini menggunakan tiga kelas atau *multiclass* maka akan menggunakan perbandingan One vs One untuk mendapatkan nilai kinerja dari *multiclass*. Dalam melakukan pemisahan SVM dibantu dengan *Kernel* yang memetakan *vector* dalam ruang dimensi yang lebih tinggi. Fungsi *kernel* seperti Linear, *Polinomial*, dan Radial (RBF). Penelitian ini menggunakan persamaan (7) RBF sebagai *kernel* SVM dengan perbandingan data secara *OvO* menggunakan *library sklearn* menggunakan fungsi SVC.

$$\text{Kernel RBF} = \exp(-\gamma(x' * y) + r) \quad (7)$$

4) *Nearest Neighbor*: merupakan metode klasifikasi objek dengan melakukan pemilihan kelas terhadap data berdasarkan jarak terdekatnya. NN merupakan algoritma yang tidak membangun model ketika melakukan klasifikasi. NN melakukan klasifikasi dengan menghitung nilai tetangga atau K menggunakan perhitungan jarak *Euclidean* pada persamaan (8). Pembangunan model NN pada penelitian ini, dibangun dengan nilai K sebanyak 60

menggunakan fungsi *KNeighborsClassifier* pada *library Sk-Learn*. Nilai K yang digunakan merupakan nilai paling optimal dari pencarian nilai K 1-100.

$$\text{Euclidian} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (8)$$

F. Evaluasi

29 Dalam melakukan evaluasi terdapat beberapa algoritma yang digunakan yaitu fungsi *fitness*, *K-Fold Cross Validation*, dan *Confusion Matrix*. Fungsi *fitness* digunakan untuk mengevaluasi nilai solusi dari setiap percobaan *metaheuristik* atau menghitung nilai dari partikel pada PSO dan kromosom pada GA. *K-Fold* digunakan untuk membagi data set menjadi data latih dan data uji. *Confusion Matrix* untuk mengevaluasi kinerja model dalam melakukan klasifikasi.

1) *Fungsi Fitness*: Fungsi ini untuk melakukan evaluasi pada algoritma *metaheuristik*. Dengan menghitung nilai dari setiap solusi pada partikel dan kromosom untuk dibandingkan antar partikel dan antar kromosom sehingga menghasilkan nilai terbaik. Nilai terbaik merupakan nilai yang mengecilkan jumlah fitur yang digunakan dengan akurasi yang lebih tinggi [21]. Hasil evaluasi *fitness* menggunakan persamaan (9) dari partikel dan kromosom akan digunakan untuk memilih fitur pada data ekstraksi TF-IDF.

$$\text{Fitness} = \alpha x(1 - P) + (1 - \alpha) x \frac{N_{\text{selection}}}{N_{\text{features}}} \quad (9)$$

30 2) *K-Fold Cross Validation*: *K-Fold* merupakan metode pembagian data set yang melakukan pemisahan data set dengan jumlah sama sebanyak K dan secara bergantian akan memilih data latih dan data uji. Hasil prediksi akan dibandingkan dengan data label / kelas untuk dievaluasi menggunakan *confusion matrix*.

3) *Confusion Matrix*: metode ini membandingkan kumpulan nilai label dengan nilai prediksi akan mendapatkan nilai kebenarannya atau kesamaan hasil. Dalam *confusion matrix* (lihat Tabel IX) dapat dihitung nilai metrik untuk mengevaluasi model percobaan prediksi mesin.

TABEL IX
CONFUSION MATRIX

| Model | Prediksi Positif | Prediksi Negatif |
|----------------|------------------|------------------|
| Aktual Positif | True Positive | False Negative |
| Aktual Negatif | False Positive | True Negative |

$$\text{Accuracy} = \frac{\text{jumlah prediksi benar}}{\text{total prediksi}} \quad (10)$$

$$\text{Precision} = \frac{\text{jumlah prediksi benar pada kelas}}{\text{total yang diprediksi sebagai kelas}} \quad (11)$$

$$\text{Recall} = \frac{\text{jumlah prediksi benar pada kelas}}{\text{total yang dipisahkan dari kelas}} \quad (12)$$

$$\text{F1-score} = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (13)$$

IV. HASIL DAN PEMBAHASAN

Dalam studi ini melakukan percobaan dengan melakukan beberapa tahapan yaitu, mengumpulkan data, *preprocessing*, *feature extraction*, *feature selection*, dan membangun model. *Feature selection* menggunakan algoritma *metaheuristik Particle Swarm Optimization* dan *Genetic Algorithm* yang menghasilkan fitur-fitur yang digunakan pada model klasifikasi. Fitur-fitur dipilih dari data set yang telah dilakukan ekstraksi fitur TF-IDF. Terdapat empat model yang akan dibangun, yaitu *Multinomial Naïve Bayes*, *Complement Naïve Bayes*, *Support Vector Machine*, dan *Nearest Neighbor*. Data set pada setiap model akan dibagi menggunakan *10-Fold Cross Validation*. Hasil prediksi setiap model akan dievaluasi menggunakan *confusion matrix* untuk mendapatkan nilai *accuracy*, *precision*, *recall*, dan *f1-score*.

Particle Swarm Optimization dibangun dengan parameter populasi 5, iterasi 10, batas atas 1, batas bawah 0, C1 0.5, C2 0.5, dan w 0.9. Sedangkan *Genetic Algorithm* dibangun dengan parameter populasi kromosom 10, generasi 10, kombinasi rate 0.9, dan mutasi rate 1/total fitur. Model *Support Vector Machine* akan dibangun dengan menggunakan *kernel Radial Basis Function* atau RBF dengan parameter gamma 0.3 dan C 0.5. Model *Nearest Neighbors* akan menggunakan nilai K sebanyak 60 yang didapatkan dari pemilihan akurasi Terbaik pada pencarian nilai k dari 1-100.

A. Hasil Implementasi Preprocessing Data

Setelah data set diberikan label positif, negatif, dan netral secara manual, kemudian dilakukan tahapan *preprocessing* yang terdiri dari tahap *case folding*, *cleaning*, *tokenisasi*, normalisasi dan *stemming* untuk menjadikan data tersebut sesuai standar kebutuhan. Hasil dari satu sampel data set yang dilakukan keseluruhan tahapan *preprocessing* ditampilkan dalam bentuk tabel yang terlihat pada Tabel X.

TABEL X
SAMPEL HASIL PREPROCESSING DATA

| Text | Tahap Preprocessing |
|--|---------------------|
| 2 Kalo ada kata korban berarti memang perlu dong adanya Permendikbud 30 2021 ini. Jelas permendikbud ini ada untuk mencegah kekerasan seksual di lingkungan kampus. Kalo sedini mungkin dicegah berarti bisa meminimalisir ada ny korban #dukungpermendikbud #sahkanpermendikbud | Tanpa Preprocessing |
| kalo ada kata korban berarti memang perlu dong adanya permendikbud 30 | Case Folding |

| Text | Tahap Preprocessing |
|---|---------------------|
| 2021 ini jelas permendikbud ini ada untuk mencegah kekerasan seksual di lingkungan kampus kalo sedini mungkin dicegah berarti bisa meminimalisir ada ny korban #dukunpermendikbud #sahkanpermendikbud | |
| kalo ada kata korban berarti memang perlu dong adanya permendikbud ini jelas permendikbud ini ada untuk mencegah kekerasan seksual di lingkungan kampus kalo sedini mungkin dicegah berarti bisa meminimalisir ada ny korban | Cleaning |
| kalau ada kata korban berarti memang perlu dong adanya permendikbud ini jelas permendikbud ini ada untuk mencegah kekerasan seksual di lingkungan kampus kalau sedini mungkin dicegah berarti bisa meminimalisir ada nya korban | Normalisasi |
| kata korban berarti permendikbud mencegah kekerasan seksual lingkungan kampus sedini dicegah meminimalisir korban | Stopword |
| kata korban arti permendikbud cegah keras seksual lingkung kampus dini cegah meminimalisir korban | Stemming |

B. Hasil Implementasi Feature Extraction TF-IDF

Data set yang melewati proses tahapan preprocessing, kemudian dilakukan ekstraksi fitur dengan menghitung nilai setiap bobot term atau kata supaya dapat dilakukan klasifikasi pada model. Penggunaan term akan menjadi fitur yang akan dihitung dengan pembobotan TF-IDF. Hasil ekstraksi fitur menggunakan pembobotan TF-IDF dapat dilihat pada Gambar 2.

| | |
|--------------------------------|-------------------------------------|
| (0, 2047) @.12199817798259122 | (15635, 5620) @.20251755837227898 |
| (0, 3148) @.42391961927755584 | (15635, 2188) @.1974066874076066 |
| (0, 9401) @.3049433997244895 | (15635, 5797) @.22234721319288426 |
| (0, 9275) @.42391961927755584 | (15635, 4151) @.193955311805482 |
| (0, 5552) @.3087888936185871 | (15635, 2898) @.5515886368823046 |
| (0, 12857) @.42391961927755584 | (15635, 18607) @.26778571759587935 |
| (0, 9817) @.16325936843284685 | (15635, 1837) @.1663617356651328 |
| (0, 18518) @.18608422542656272 | (15635, 12342) @.461450383723483264 |
| (0, 9458) @.0508645988515751 | (15635, 5598) @.18781682044653523 |
| (0, 10643) @.211687692145304 | (15635, 2947) @.19527383318258482 |
| (0, 2344) @.27851654858798596 | (15635, 9458) @.84078767737954366 |
| (0, 6834) @.19582991819678434 | (15636, 39) @.5434822216463286 |
| (1, 10948) @.456879371177347 | (15636, 3925) @.46137759182812527 |
| (1, 5681) @.4768825384429885 | (15636, 3489) @.38282769451397 |
| (1, 1797) @.713726364576794 | (15636, 4948) @.28632361462088067 |
| (1, 9458) @.23494964081831797 | (15636, 1692) @.2579788961815836 |
| (2, 4589) @.3041289912873039 | (15636, 7364) @.2593886236913399 |
| (2, 5407) @.334873873458893 | (15636, 12458) @.1782957245811803 |
| (2, 6996) @.2808963732822154 | (15636, 9655) @.20589273858281713 |
| (2, 7259) @.3328643346080974 | (15636, 18607) @.2144542199156767 |
| (2, 10965) @.2832584724786646 | (15636, 9458) @.13840177575128212 |

Gambar 2 Hasil ekstraksi fitur TF-IDF

C. Hasil Implementasi Model Machine Learning

Data set yang telah dilakukan ekstraksi fitur akan dibagi menggunakan 10-Fold Cross Validation. Selanjutnya akan dibangun model Multinomial Naive Bayes, Support Vector Machine, Complement Naive Bayes, dan Nearest Neighbor dan dievaluasi menggunakan standar deviasi nilai setiap metrik di confusion matrix setiap nilai foldnya. Hasil standar deviasi metrik accuracy,

precision, recall, dan f1-score dari evaluasi 10-fold cross validation pada model machine learning terdapat pada Tabel XI.

TABEL XI
STANDAR DEVIASI K-FOLD CROSS VALIDATION

| Model | Accuracy | Precision | Recall | F1-Score |
|-------|----------|-----------|--------|----------|
| MNB | 0.072 | 0.059 | 0.072 | 0.085 |
| CNB | 0.072 | 0.072 | 0.072 | 0.075 |
| SVM | 0.088 | 0.068 | 0.088 | 0.102 |
| KNN | 0.073 | 0.068 | 0.073 | 0.081 |

Kemudian akan dibandingkan kinerja keseluruhan model dalam melakukan prediksi kelas positif, negatif, dan netral menggunakan nilai dari metrik f1-score. Hasil perbandingan kemampuan klasifikasi model terdapat pada Tabel XII.

TABEL XII
HASIL F1-SCORE PADA SETIAP KELAS

| Model | Positif | Negatif | Netral |
|-------|---------|---------|--------|
| MNB | 0.745 | 0.620 | 0.260 |
| CNB | 0.754 | 0.673 | 0.518 |
| SVM | 0.751 | 0.640 | 0.414 |
| KNN | 0.751 | 0.650 | 0.434 |

Membandingkan nilai accuracy, dan nilai rerata weighted dari precision, recall, dan f1-score yang didapatkan dari evaluasi confusion matrix pada setiap model machine learning. Hasil perbandingan ditampilkan dalam bentuk tabel yang terlihat pada Tabel XIII.

TABEL XIII
HASIL CONFUSION MATRIX SETIAP MODEL

| Model | Accuracy | Precision | Recall | F1-Score |
|-------|----------|-----------|--------|----------|
| MNB | 0.659 | 0.734 | 0.659 | 0.611 |
| CNB | 0.688 | 0.685 | 0.688 | 0.683 |
| SVM | 0.678 | 0.728 | 0.678 | 0.651 |
| KNN | 0.680 | 0.702 | 0.680 | 0.657 |

D. Hasil Evaluasi Feature Selection Particle Swarm Optimization

Particle Swarm Optimization dengan evaluasi fitness menggunakan nilai alfa 0.9 dan P merupakan akurasi dari model machine learning pada setiap partikel. Partikel yang memiliki nilai fitness paling sedikit dengan jumlah fitur sedikit dan akurasi yang tinggi dalam semua pencarian akan menjadi Global Best. Posisi partikel pada global best akan menggunakan data ekstraksi fitur TF-IDF yang sudah dibagi menggunakan 10-fold cross validation dan dievaluasi menggunakan confusion matrix. Kinerja model akan dievaluasi menggunakan metrik accuracy, rerata weighted precision, rerata weighted recall, dan rerata weighted f1-score. Hasil percobaan ditampilkan dalam bentuk tabel pada Tabel XIV.

TABEL XIV
PERBANDINGAN FEATURE SELECTION PSO PADA SETIAP MODEL

| Model | Accuracy | Precision | Recall | F1-Score |
|-------|----------|-----------|--------|----------|
| MNB | 0.667 | 0.713 | 0.667 | 0.628 |
| CNB | 0.689 | 0.686 | 0.689 | 0.684 |

10

| Model | Accuracy | Precision | Recall | F1-Score |
|-------|----------|--------------|--------|----------|
| SVM | 0.662 | 0.728 | 0.662 | 0.626 |
| KNN | 0.678 | 0.703 | 0.678 | 0.655 |

E. Hasil Evaluasi Feature Selection Particle Genetic Algorithm

Genetic Algorithm dengan evaluasi fitness menggunakan nilai *alpha* 0.9 dan P merupakan akurasi dari setiap model machine learning pada setiap kromosom. Kromosom yang mempunyai nilai fitness yang lebih baik dengan fitur sedikit dan akurasi tinggi. Kromosom terbaik akan dievaluasi menggunakan data ekstraksi fitur TF-IDF yang dibagi dengan 10-fold cross validation dan kinerja model akan dievaluasi menggunakan confusion matrix. Metrik yang akan digunakan yaitu accuracy, rerata weighted precision, rerata weighted recall, dan rerata weighted f1-score. Hasil percobaan feature selection menggunakan genetic algorithm ditampilkan pada Tabel XV.

TABEL XV
PERBANDINGAN FEATURE SELECTION GA PADA SETIAP MODEL

| Model | Accuracy | Precision | Recall | F1-Score |
|-------|--------------|--------------|--------------|--------------|
| MNB | 0.632 | 0.727 | 0.632 | 0.569 |
| CNB | 0.666 | 0.663 | 0.666 | 0.663 |
| SVM | 0.638 | 0.714 | 0.638 | 0.584 |
| KNN | 0.624 | 0.655 | 0.624 | 0.586 |

F. Perbandingan Keseluruhan Percobaan

Hasil percobaan yang telah dilakukan akan dilakukan perbandingan feature selection dalam meningkatkan kinerja model machine learning. Membandingkan percobaan tanpa feature selection atau menggunakan semua fitur, feature selection particle swarm optimization, dan feature selection genetic algorithm pada setiap model machine learning yang ditampilkan pada Tabel XVI

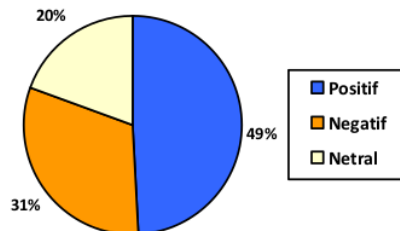
TABEL XVI
PERBANDINGAN KESELURUHAN PERCOBAAN

| Model | Seleksi Fitur | Accuracy | Precision | Recall | F1-Score |
|-------|---------------|--------------|-----------|--------|----------|
| MNB | tanpa | 0.659 | 0.734 | 0.659 | 0.611 |
| | PSO | 0.667 | 0.713 | 0.667 | 0.628 |
| | GA | 0.632 | 0.727 | 0.632 | 0.569 |
| CNB | tanpa | 0.688 | 0.685 | 0.688 | 0.683 |
| | PSO | 0.689 | 0.686 | 0.689 | 0.684 |
| | GA | 0.666 | 0.663 | 0.666 | 0.663 |
| SVM | tanpa | 0.678 | 0.728 | 0.678 | 0.651 |
| | PSO | 0.662 | 0.728 | 0.662 | 0.626 |
| | GA | 0.638 | 0.714 | 0.638 | 0.584 |
| KNN | tanpa | 0.680 | 0.702 | 0.680 | 0.657 |
| | PSO | 0.678 | 0.703 | 0.678 | 0.655 |
| | GA | 0.624 | 0.655 | 0.624 | 0.586 |

G. Analisis Keseluruhan Percobaan

Pengambilan data yang dilakukan dengan 3 kata kunci yaitu permendikbud, #dukunpermendikbud, dan #cabutpermendikbud menghasilkan data bersih sebanyak 15.637 data. Dari data tersebut dilakukan proses labelisasi

dengan memberikan label atau kelas positif, negatif, dan netral pada data set secara manual. Hasil proses labelisasi dapat dilihat pada Gambar 3.



Gambar 3 Sentimen permendikbudristek PPKS-LPT

Berdasarkan proses tersebut mayoritas masyarakat mendukung adanya dan pengesahan Permendikbudristek PPKS-LPT. Akan tetapi, gabungan responden negatif dan netral mempunyai jumlah yang lebih besar dibandingkan responden positif sehingga dukungan Permendikbudristek PPKS-LPT ini tidak absolut, sehingga masih ada kemungkinan masyarakat berubah pikiran. Berdasarkan responden yang didapatkan pendukung merupakan orang-orang yang memiliki ketertarikan dengan kasus kekerasan seksual atau orang-orang yang berempati pada korban. Sedangkan orang yang menolak Permendikbudristek 30 tahun 2021 merupakan orang-orang yang takut terjadinya pelecehan perzinahan di lingkungan pendidikan dan ketakutan akan penggunaan frasa “tanpa persetujuan korban” akan melegalkan pelaku jika sudah mendapatkan persetujuan dari korban sehingga menjadi perzinahan. Responden yang mengambil sikap netral merupakan orang-orang yang menimbang keunggulan dan kekurangan dalam keberpihakan serta menuliskan pandangan secara ambigu yang keberpihakkannya diserahkan kepada pembaca.

Empat model machine learning yang akan digunakan dalam proses percobaan klasifikasi dilakukan evaluasi kemampuan mengenali model dalam setiap kelasnya. Evaluasi pertama dengan mengambil nilai standar deviasi setiap model yang terlihat pada Tabel XI. Dalam melakukan 10-fold cross validation menghasilkan model prediksi setiap foldnya. Model prediksi tersebut dilakukan evaluasi menggunakan confusion matrix. Hasilnya akan terdapat 10 nilai evaluasi confusion matrix yang akan dicari nilai standar deviasi untuk mengetahui kemampuan model dalam menggeneralisasi klasifikasi pada data set yang berbeda-beda. Dari Tabel XI, diketahui model Complement Naïve Bayes memiliki kinerja klasifikasi paling baik dibandingkan model lainnya dibuktikan dengan nilai standar deviasi setiap matriksnya sangat berbeda dibandingkan model lainnya. Dengan nilai f1-score pada prediksi model mengenali setiap kelas yang ditampilkan pada Tabel XII, kemampuan model Multinomial Naïve Bayes memiliki kinerja paling buruk dengan kegagalan mengenali kelas netral, sedangkan

model CNB, SVM, dan NN mampu mengenali kelas lain dengan lebih baik, dengan model CNB sebagai model paling baik dalam mengenali label positif, negatif, dan netral. Mengambil nilai *confusion matrix* pada setiap model sebagai model tanpa *feature selection* atau menggunakan keseluruhan fitur yang ditampilkan pada Tabel XIII, CNB memiliki kinerja paling baik dibandingkan yang lain walaupun *precision* tidak lebih baik dibandingkan MNB, akan tetapi karena CNB memiliki kinerja tertinggi pada 3 metrik maka CNB merupakan model paling baik.

Pencarian fitur *Particle Swarm Optimization* menghasilkan *Global Best* dengan nilai terbaik pada posisi partikel sebagai penggunaan fitur-fitur pada data ekstraksi fitur TF-IDF. Pencarian PSO mampu mendapatkan fitur-fitur terpilih yang terbaik dengan jumlah fitur digunakan sedikit dengan akurasi yang tinggi. Dalam pencarian fitur pada *feature selection* menggunakan PSO yang ditampilkan pada Tabel XIV menghasilkan kinerja model CNB yang lebih baik dibandingkan kinerja model MNB, SVM, dan KNN. Dengan nilai *accuracy* CNB sebanyak 0.689, *precision* 0.686, *recall* 0.689, dan *f1-score* 0.684. Kemampuan *feature selection* PSO dalam mencari fitur terbaik mampu memilih fitur-fitur terbaik yang meningkatkan kinerja model klasifikasi.

Pencarian fitur *Genetic Algorithm* menghasilkan nilai *fitness* terbaik pada populasi kromosom. Nilai gen pada kromosom digunakan dengan sebagai fitur terpilih pada data ekstraksi fitur TF-IDF. Hasil fitur terpilih dari kromosom terbaik akan dievaluasi menggunakan *confusion matrix* pada setiap model yang ditampilkan pada Tabel XV mampu menghasilkan kinerja yang baik walaupun menggunakan fitur yang sedikit. Model CNB menjadi model dengan kinerja paling tinggi. Kemampuan keempat model pada fitur terpilih *genetic algorithm* tidak berbeda jauh dengan perbedaan kurang dari 5%.

Membandingkan keseluruhan percobaan yang telah dilakukan dengan menggunakan semua fitur maupun dengan menggunakan fitur terpilih dari *feature selection* yang sudah dirangkum pada Tabel XVI. Menggunakan fitur terpilih dari *feature selection* PSO mampu meningkatkan kinerja akurasi pada model CNB sebanyak 0.1% dan MNB sebanyak 0.8%. Sedangkan pada model SVM mengalami penurunan sebanyak 1.6% dan pada model KNN turun sebanyak 0.2%. Kemampuan *feature selection* PSO meningkatkan CNB sesuai dengan kemampuan model CNB yang memiliki model dengan kinerja seimbang dan paling baik dibandingkan yang lain. Model CNB sendiri, mampu bekerja optimal pada data yang tidak seimbang sedangkan model lain masih kurang optimal untuk data tidak seimbang. Menggunakan fitur terpilih dari *feature selection* GA masih belum mampu mendapatkan akurasi yang lebih baik dibandingkan menggunakan semua fitur dengan penurunan akurasi paling tinggi pada SVM sebanyak 4.6%. Secara keseluruhan penggunaan *feature selection* mampu menghasilkan kinerja yang mendekati atau lebih baik dibandingkan dengan menggunakan semua fitur, selain itu

penggunaan fitur yang lebih sedikit akan mengurangi waktu komputasi yang dibutuhkan dalam melakukan klasifikasi.

V. KESIMPULAN

7) Algoritma *Multinomial Naïve Bayes*, *Complement Naïve Bayes*, *Support Vector Machine*, dan *Nearest Neighbor* dapat memprediksi kelas sentimen Permendikbudristek PPKS dari data teks media sosial Twitter. Dari hasil perbandingan keempatnya, algoritma CNB memiliki kinerja paling baik dengan nilai *accuracy* 0.688, *precision* 0.685, *recall* 0.688, dan *f1-score* 0.683.

Algoritma *Genetic Algorithm* dan *Particle Swarm Optimization* dapat digunakan dalam *feature selection* data dari ekstraksi fitur TF-IDF. GA masih belum mampu menemukan kombinasi fitur yang memiliki kinerja lebih baik dibandingkan dengan menggunakan semua fitur. PSO berhasil menemukan kombinasi fitur yang memiliki kinerja sama atau lebih baik dibandingkan dengan menggunakan semua fitur. Peningkatan akurasi pada PSO paling tinggi sebanyak 0,8% pada model *Multinomial Naïve Bayes*.

Sentimen Masyarakat secara umum berada pada kelas positif sebanyak 49%, kelas negatif sebanyak 31%, dan kelas netral sebanyak 20%.

REFERENSI

- [1] J. Kenedi, *Perlindungan Saksi dan Korban (Studi Perlindungan Hukum dan Kejahatan dan Sistem Peradilan di Indonesia)*. 1st ed. Yogyakarta: PUSTAKA BELAJAR
- [2] Komnas Perempuan, "Lembar Fakta Kekerasan Seksual di Lingkungan Pendidikan," *Komisi Nasional Anti Kekerasan terhadap Perempuan*, Indonesia, 2020
- [3] KPAI, "SIARAN PERS : Catatan Pelanggaran Anak Tahun 2021 dan Proyeksi Pengawasan Penyelenggaraan Perlindungan Anak Tahun 2022," *Komisi Perlindungan Anak Indonesia*, Indonesia, 2021.
- [4] Dirjen DIKTI, "Naskah Akademik Pendukung Urgensi Draft Peraturan Menteri tentang Pencegahan dan Penanggulangan Kekerasan Seksual di Lingkungan Perguruan Tinggi," *Direktorat Jenderal Pendidikan Tinggi Kementerian Pendidikan dan Kebudayaan Republik Indonesia*, Indonesia, 2020.
- [5] Kemendikbud, "Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 30 Tahun 2021 tentang Pencegahan dan Penanganan Kekerasan Seksual di Lingkungan Perguruan Tinggi," *Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia*, Jakarta, 2021.
- [6] We Are Sosial, "Digital 2021: Indonesia Digital Report 2021," *HootSuite*, London, 2021.
- [7] L. P. Hung, M. H. A. Hijazi, and R. Alfred, "A Review on Feature selection Methods for Sentimen Analysis". *J. Computational and Theoretical Nanoscience*, vol 21, pp. 2957-2962, 2015
- [8] S. R. Hakim, M. A. Rizki, N. I. Zekha, N. Fitri, Y. R. A. and R. Nooraeni, "Analisis sentimen pengguna Instagram terhadap kebijakan kemdikbud mengenai bantuan kuota internet dengan metode *support vector machine* (SVM)," *Jurnal Matematika dan Statistika serta Aplikasinya*, vol. 8, Des. 2020.
- [9] G. D. Hamidi, F. A. Bestari, A. Situmorang, and N. A. Rakhmawati, "Sentiment Analysis on the Ratification of Penghapusan Kekerasan Seksual Bill on Twitter." *JuTISI : J. Teknik Informatika dan Sistem Informasi*, vol. 7, no.3, Des. 2021
- [10] Yennimar and R. A. Rizal, "Comparison of Machine learning Classification Algorithms in Sentiment Analysis Product Review of North Padang Lawas Regency," *SinkOn : Journal*

- Publications and Informatics Engineering Research*, vol. 4, no. 1, pp. 268-273, 2019.
- [11] S. Ernawati, and R. Wati, "Penerapan Algoritma K-Nearest Neighbors pada Analisis Sentimen Review Agen Trave". *J. Khaulistiwa Informatika*, vol. 6, no. 1, pp. 64-69. 2018.
- [12] A. Deviyanto, R. Wahyudi, and M. Didik, "Penerapan Analisis Sentimen pada Pengguna Twitter Menggunakan Metode K Nearest Neighbor," *JISKa : Jurnal Informatika Sunan Kalijaga*, vol. 3, no. 1, pp. 1-13, 2018.
- [13] A. R. Isnain, J. Supriyanto, and M. P. Kharisma, "Implementation of K Nearest Neighbor (K-NN) Algorithm for Public Sentiment Analysis of Online Learning," *IJCCS: Indonesian Journal Computing and Cybernetics Systems*, vol. 15, no. 2, pp. 121-130, Apr. 2021.
- [14] E. Purnamasari, D. P. Rini, and Sukemi, "Seleksi Fitur menggunakan Algoritma Particle Swarm Optimization pada Klasifikasi Kelulusan Mahasiswa dengan Metode Naive Bayes," *J. RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 3, Jul. 2020.
- [15] E. Indrayuni, and A. Nurhadi, "Optimizing Genetic Algorithms for Sentiment Analysis of Apple Product Reviews using SVM," *Sinkron : Jurnal dan Penelitian Teknik Informatika*, vol. 2, no. 4, Apr. 2020.
- [16] M. Rostami, K. Berahmand, and S. Forouzandeh, "A Novel Community Detection based Genetic Algorithm for Feature selection," *J. Big Data*, vol. 8, no. 2, pp. 1-27, 2021.
- [17] A. Garcia-Dominguez, C. E. Galvan-Tejada, L. A. Zanella-Calzada, H. Gamboa-Rosales, J. I. Galvan-Tejada, J. M. Celaya-Padilla, H. Luna-Garcia, and R. Mallanes-Quintanar, "Feature selection Using Genetic Algorithms for the Generation of a Recognition and Classification of Children Activities Model Using Environmental Sound", *Mobile Information System*, vol. 2020, no. 8617430, pp. 12, 2020.
- [18] K. J. Prayoga, A. Nugroho, and T. N. Wiyatno, "Komparasi Feature selection Particle Swarm Optimization (PSO) dengan Genetic Algorithm (GA) terhadap Algoritma Naive Bayes pada Analisis Sentimen Twitter," *Proc. SNasTekS*, Bekasi, 2019.
- [19] M. Avinash, and E. Sivansankar, "Efficient Feature selection Techniques for Sentiment Analysis," pp. 11911.00288v2, Feb. 2020.
- [20] S. M. Tambunan, Y. Nataliani, and E. S. Lestari, "Perbandingan Klasifikasi dengan Pendekatan Pembelajaran Mesin untuk Mengidentifikasi Tweet Hoaks di Media Sosial Twitter," *J. JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 7, no. 2, pp. 112-120, Aug. 2021.
- [21] S. M. Viera, L. F. Mendonca, G. J. Fainha, and J. M. Sousa, "Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients," *Applied Soft Computing*, vol. 13, no. 8, pp 3494-3504, Aug. 2013.

Artikel

ORIGINALITY REPORT

18%

SIMILARITY INDEX

16%

INTERNET SOURCES

11%

PUBLICATIONS

5%

STUDENT PAPERS

PRIMARY SOURCES

| | | |
|---|--|-----|
| 1 | jurnal.untan.ac.id Internet Source | 4% |
| 2 | github.com Internet Source | 3% |
| 3 | journal.trunojoyo.ac.id Internet Source | 1% |
| 4 | channel9.id Internet Source | 1% |
| 5 | Submitted to University of Greenwich Student Paper | 1% |
| 6 | dspace.uui.ac.id Internet Source | <1% |
| 7 | repository.bsi.ac.id Internet Source | <1% |
| 8 | Dewi Mardawati, Linda Handayuni, Ririn Afrima Yenni, Claudya Rizki Audina Daulay, Kevin Wahyudi Farhan, Nidno Friani Candra. "Sosialisasi Pentingnya Penggunaan Masker pada Pembelajaran Tatap Muka pada Masa | <1% |

Pandemi COVID-19 di Sekolah Menengah Atas (SMA) Kota Padang", Jurnal Surya Medika, 2022

Publication

9

repositor.umm.ac.id

Internet Source

<1 %

10

Ivan Rifky Hendrawan, Ema Utami, Anggit Dwi Hartanto. "Comparison of Naive Bayes Algorithm and XGBoost on Local Product Review Text Classification", Edumatic: Jurnal Pendidikan Informatika, 2022

Publication

<1 %

11

Rinna Rachmatika, Achmad Bisri. "Perbandingan Model Klasifikasi untuk Evaluasi Kinerja Akademik Mahasiswa", Jurnal Edukasi dan Penelitian Informatika (JEPIN), 2020

Publication

<1 %

12

ecampus.pelitabangsa.ac.id

Internet Source

<1 %

13

text-id.123dok.com

Internet Source

<1 %

14

www.sciencegate.app

Internet Source

<1 %

15

jurnal.iaii.or.id

Internet Source

<1 %

16

j-ptiik.ub.ac.id

Internet Source

<1 %

17

Submitted to UIN Sultan Syarif Kasim Riau

Student Paper

<1 %

18

Submitted to Universitas Diponegoro

Student Paper

<1 %

19

ejournal.upbatam.ac.id

Internet Source

<1 %

20

www.scribd.com

Internet Source

<1 %

21

Maulidya Dwi Nurmalasari, Kusrini Kusrini, Sudarmawan Sudarmawan. "Komparasi Algoritma Naive Bayes dan K-Nearest Neighbor untuk Membangun Pengetahuan Diagnosa Penyakit Diabetes", Jurnal Komtika (Komputasi dan Informatika), 2021

Publication

<1 %

22

Sri Juniarsih, Eva Faja Ripanti, Enda Eisyudha Pratama. "Implementasi Naive Bayes Classifier pada Opinion Mining Berdasarkan Tweets Masyarakat Terkait Kinerja Presiden dalam Aspek Ekonomi", Jurnal Sistem dan Teknologi Informasi (Justin), 2020

Publication

<1 %

23

repository.unhas.ac.id

Internet Source

<1 %

| | | |
|----|--|------|
| 24 | <p>Giat Subroto, Nina Sulistiyowati, Azhari Ali Ridha. "Klasifikasi Jenis Kekerasan Pada Perempuan Dan Anak Dengan Algoritma Multinomial Naïve Bayes", INTECOMS: Journal of Information Technology and Computer Science, 2022</p> <p>Publication</p> | <1 % |
| 25 | <p>Yashoda Barve, Jatinderkumar R. Saini. "Detecting and Classifying Online Health Misinformation with 'Content Similarity Measure (CSM)' algorithm: An Automated Fact-checking based Approach", Research Square Platform LLC, 2022</p> <p>Publication</p> | <1 % |
| 26 | <p>amincorporatio.blogspot.com</p> <p>Internet Source</p> | <1 % |
| 27 | <p>docplayer.info</p> <p>Internet Source</p> | <1 % |
| 28 | <p>repository.ub.ac.id</p> <p>Internet Source</p> | <1 % |
| 29 | <p>123dok.com</p> <p>Internet Source</p> | <1 % |
| 30 | <p>core.ac.uk</p> <p>Internet Source</p> | <1 % |
| 31 | <p>digilib.uin-suka.ac.id</p> <p>Internet Source</p> | <1 % |

| | | |
|----|--|------|
| 32 | faculty.petra.ac.id Internet Source | <1 % |
| 33 | index.pkp.sfu.ca Internet Source | <1 % |
| 34 | journal.unisnu.ac.id Internet Source | <1 % |
| 35 | media.neliti.com Internet Source | <1 % |
| 36 | softscients.com Internet Source | <1 % |
| 37 | Ahmad Fathan Hidayatullah, Aufa Aulia Fadila, Kiki Purnama Juwairi, Royan Abida Nayoan. Jurnal Linguistik Komputasional (JLK), 2019 Publication | <1 % |
| 38 | Yunita Maulidia Sari, Nenden Siti Fatonah. "Peringkasan Teks Otomatis pada Modul Pembelajaran Berbahasa Indonesia Menggunakan Metode Cross Latent Semantic Analysis (CLSA)", Jurnal Edukasi dan Penelitian Informatika (JEPIN), 2021 Publication | <1 % |
| 39 | ejournal.unsri.ac.id Internet Source | <1 % |

Exclude quotes On

Exclude matches Off

Exclude bibliography On

Artikel

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10
