Jurnal Manajemen Informatika (JAMIKA)
Volume 13 Nomor 1 Edisi April 2023
E ISSN: 2655-6960 | P ISSN: 2088-4125
OJS: https://ojs.unikom.ac.id/index.php/jamika

Deteksi Komentar *Cyberbullying* pada Media Sosial *Instagram* Menggunakan Algoritma *Random Forest*

Cyberbullying Comment Detection on Instagram Social Media Using Random Forest Algorithm

Heri Santoso[1], **Raissa Amanda Putri**[2], **Sahbandi**[3*]
Program Studi Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
*E-mail: sahbandi112237@gmail.com

Abstrak

Cyberbullying adalah fenomena di media sosial dengan menggunakan perangkat teknologi untuk menghina, merendahkan, dan tidak menghargai orang lain. Hal ini dapat mengakibatkan gangguan mental seperti kehilangan rasa percaya diri, stres, depresi, bahkan dorongan untuk bunuh diri. Survei Ditch The Label lembaga riset Inggris menetapkan Instagram sebagai media sosial dengan jumlah cyberbullying tertinggi. Tujuan penelitian ini adalah untuk mengetahui akurasi terbaik berdasarkan hasil klasifikasi dataset komentar cyberbullying serta mendeteksi komentar baru apakah termasuk dalam kelas bullying atau non-bullying. Salah satu metode yang dapat digunakan adalah algoritma random forest yang menggabungkan beberapa metode sejenis ataupun berbeda seperti decision tree pada proses klasifikasi. Hasil klasifikasi pada data testing menggunakan algoritma random forest menunjukkan akurasi terbaik sebesar 84% pada kombinasi hyperparameteres tuning terakhir. Model yang dibangun juga dapat mendeteksi komentar baru dengan hasil prediksi yang cukup baik. Saran untuk penelitian selanjutnya dapat melakukan klasifikasi komentar cyberbullying dengan kelas yang lebih spesifik, seperti komentar rasisme atau seksisme.

Kata kunci: *Cyberbullying, Instagram, Klasifikasi, Random Forest.*

Abstract

Cyberbullying is a phenomenon on social media where technological devices are used to insult, demean, and disrespect others. This can cause mental disorders such as loss of self-confidence, stress, depression, and even suicidal tendencies. The Ditch The Label survey, conducted by a British research institute, identified Instagram as the social media platform with the highest incidence of cyberbullying. The aim of this research is to determine the best accuracy based on the classification results of the cyberbullying comment dataset and to detect new comments as either bullying or non-bullying. One method that can be used is the random forest algorithm, which combines several similar or different methods, such as decision trees, in the classification process. The results of the classification of the testing data using the random forest algorithm show the highest accuracy of 84% in the last hyperparameter tuning combination. The built model can also detect new comments with fairly good predictive results. Suggestions for further research include classifying cyberbullying comments into more specific categories, such as racist or sexist comments.

Keywords: *Cyberbullying, Instagram, Classification, Random Forest.*

Naskah diterima 12 Feb. 2023; direvisi 21 Mar. 2023; dipublikasikan 1 Apr. 2023.

I. PENDAHULUAN

Berdasarkan laporan Asosiasi Penyelenggara Jasa Internet Indonesia [1], populasi penduduk Indonesia mencapai 266,91 juta jiwa. Lebih dari 70% atau sekitar 196,71 juta orang terhubung dengan jaringan internet sepanjang tahun 2019-2020, seiring dengan perkembangan teknologi informasi, internet, dan media sosial yang membawa dampak perubahan terhadap perilaku manusia dalam berinteraksi[2]. Media sosial memberikan kemudahan ,kenyamanan serta lebih efisien digunakan untuk berkomunikasi [3]. Contoh media sosial yang banyak digunakan saat ini antara lain *Whatsapp, Tiktok, Facebook, Youtube, Twitter, Instagram*, dan lain-lain [4]. Media sosial biasa digunakan sebagai sarana komunikasi, informasi serta sebagai media hiburan bagi penggunanya [5].

Instagram merupakan salah satu *platform* media sosial yang sering digunakan oleh masyarakat terlebih lagi kalangan milenial khususnya di Indonesia. Banyak orang yang tertarik dengan keunggulan media sosial tersebut karena *instagram* menawarkan ruang untuk berbagi foto dan video serta kemampuan untuk mengaplikasikan filter digital [6]. Akan tetapi dari banyaknya manfaat yang didapatkan masih banyak penggunanya yang tidak memahami etika dalam menggunakan media sosial yang justru menjadikan media

Jurnal JAMIKA Manajemen Informatika
Jurnal Manajemen Informatika (JAMIKA)
Volume 13 Nomor 1 Edisi April 2023
E ISSN: 2655-6960 | P ISSN: 2088-4125
OJS: https://ojs.unikom.ac.id/index.php/jamika

sosial sebagai sarana untuk mengintimidasi orang lain di belakang layar. Lembaga riset asal inggris *Ditch The Label* yang fokus pada *bullying* menetapkan *Instagram* sebagai media sosial dengan kasus *cyberbullying* paling banyak [7]. Di Indonesia diatur dalam Pasal 27 (3) Undang-Undang Informasi Elektronik Nomor 11 Tahun 2008, yang meliputi penyebaran atau pencemaran nama baik.

Cyberbullying adalah bentuk intimidasi untuk melecehkan orang lain menggunakan perangkat teknologi. Korban dipermalukan dengan berbagai cara oleh pelaku, termasuk melalui pesan dengan kata-kata kasar atau gambar yang mengganggu. Tidak jarang tindakan *cyberbullying* dapat menyebabkan gangguan mental terhadap korbannya apabila tidak ditangani dengan baik seperti stress, depresi, hilangnya kepercayaan diri hingga yang paling fatal munculnya dorongan untuk bunuh diri [8].

Akan tetapi tidak semua komentar yang terdapat pada media sosial merupakan *cyberbullying* begitu juga pada media sosial *instagram*. komentar tersebut bisa dibagi mana yang mengandung makna *bullying* dan mana yang *non bullying*, salah satunya adalah dengan menerapkan algoritma *machine learning* untuk melakukan klasifikasi terhadap suatu teks komentar. Penelitian ini menggunakan algoritma *Random forest* untuk mengetahui suatu teks komentar di *instagram* mengandung makna *bullying* atau *non bullying* karena merujuk pada jurnal-jurnal yang dijadikan sebagai acuan penelitian menunjukkan bahwa algoritma *Random forest* dapat memprediksi keluaran dengan tingkat akurasi yang cukup baik. *Random forest* merupakan metode yang populer digunakan untuk klasifikasi berbasis *ensemble* dan pengelompokan pohon keputusan *(decision tree)* [9]. *Random forest* merupakan pengembangan lebih lanjut dari metode *Classification and Regression Tree (CART)*, yaitu dengan menerapkan metode *bootstrap (bagging)* dan pemilihan fitur secara acak.

Adapun tujuan yang ingin dicapai pada penelitian ini, yaitu untuk mengetahui akurasi terbaik dari algoritma *Random Forest* dalam melakukan klasifikasi *dataset* komentar *cyberbullying* berbahasa Indonesia pada media sosial *instagram* serta untuk mendeteksi komentar baru apakah termasuk kedalam kelas *bullying* atau *non-bullying.* Penelitian sebelumnya yang dijadikan sebagai referensi pada penelitian ini yaitu penelitian yang dilakukan oleh Candra & Nanda Rozana [10] dengan menggunakan metode *K-Nearest Neighbor* dalam klasifikasi komentar *bullying* pada *instagram* berhasil melakukan klasifikasi komentar bully dan tidak bully, dimana akurasi terbaik yang diperoleh sebesar 77% dengan 90:10 data pada *k* 13 dan *fold* ke 6 sedangkan akurasi terendah sebesar 35% dengan 90:10 data pada *k* 11 dan *fold* ke 1. Penelitian lainnya yang dilakukan oleh Afdhal dkk [10] mengenai penerapan algoritma *random forest* untuk analisis sentimen komentar di *youtube* tentang *islamofobia* diperoleh akurasi mencapai 79% dengan *F1-Score* sebesar 86,26% yang menunjukkan bahwa algoritma *random forest* cukup mampu mengklasifikasikan sentimen pada komentar *YouTube* tentang *Islamofobia.*

Berdasarkan pemaparan permasalahan diatas, yang membedakan penelitian ini dengan penelitian terdahulu adalah belum terdapat algoritma serupa dalam mengklasifikasi, mengidentifikasi maupun mendeteksi topik yang sama, sehingga penelitian ini mengusulkan algoritma *Random Forest* untuk melakukan klasifikasi serta mendeteksi komentar *bullying* dan *non-bullying* berbahasa Indonesia pada media sosial *instagram* dan menggunakan *FastText* sebagai *word embedding*. Selain itu penelitian ini tidak hanya sebatas melakukan klasifikasi terhadap *dataset* saja, akan tetapi model yang dibangun juga dapat mendeteksi komentar baru untuk mengetahui kelas dari komentar tersebut.

II. METODE PENELITIAN

Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan dengan cara mengunjungi profil akun *Instagram* artis, selebgram, serta *public figure* Indonesia yang memiliki jumlah *follower* diatas 500 ribu dengan rentang waktu posting antara Agustus 2021 s/d April 2022 dan memilih foto atau video yang menjadi bahan penelitian. Kemudian komentar yang terdapat pada foto atau video tersebut akan di-*copy* dan disusun kedalam *Microsoft Excel* selanjutnya dilakukan pelabelan secara manual terhadap *dataset* dimana untuk setiap komentar yang memiliki label *bullying* akan diberi kelas 0 dan untuk komentar yang memiliki label *non-bullying* akan diberi kelas 1 kemudian *dataset* tersebut disimpan kedalam format *.csv (Comma Separated Values).*

TABEL 1
CONTOH *DATASET* KOMENTAR *INSTAGRAM*

No	Comment	Class
1	Cantik nya cantik nya ganteng nya ganteng nya	1
123	Sok ngartis lu, anak kemaren sore doang belagu amat lu ajg	0
781	Jerome ni sama kayak aku kalau foto pasti nyengir mulu	1
1000	AOWKAKWK NGAKAK GABISA NADA TINGGI LAGU NYA SENDIRI	0

Jurnal Manajemen Informatika (JAMIKA)
Volume 13 Nomor 1 Edisi April 2023
E ISSN: 2655-6960 | P ISSN: 2088-4125
OJS: https://ojs.unikom.ac.id/index.php/jamika

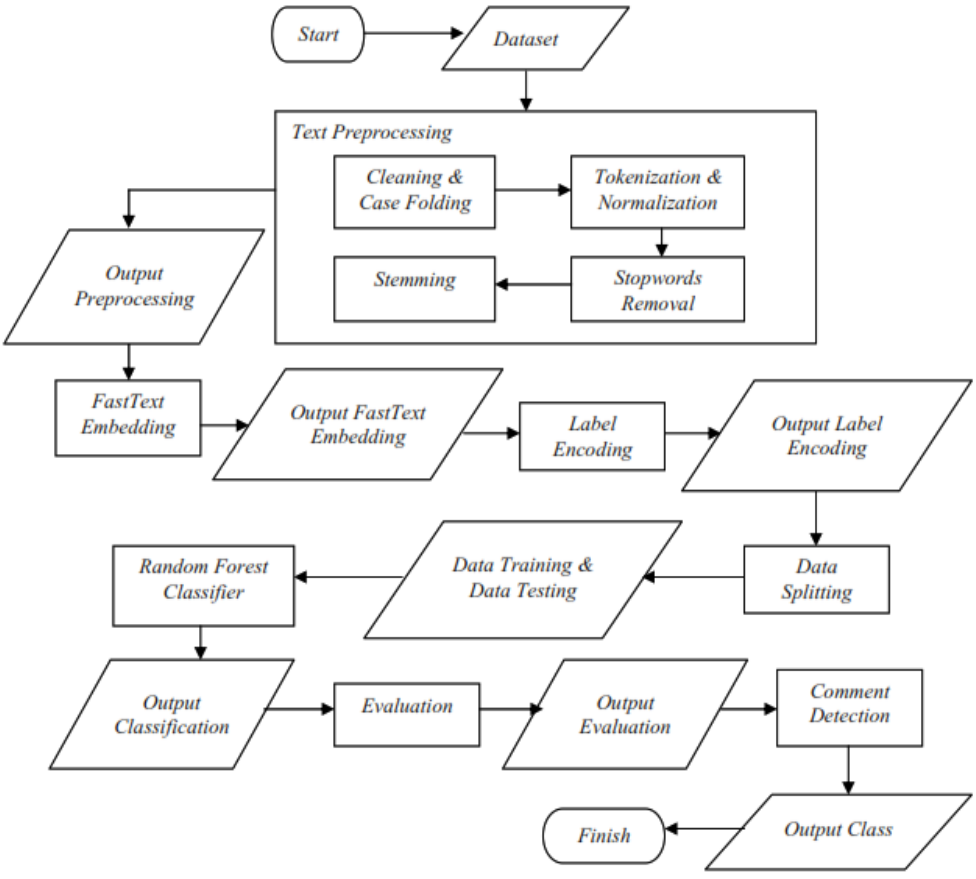
Dataset yang dikumpulkan berupa 1000 data komentar dari media sosial *instagram public figure*, artis maupun selebgram dimana terdapat 500 komentar dengan label *bullying* dan 500 komentar dengan label *non-bullying*. Pengumpulan data yang mengandung makna *bullying* didasari oleh kemunculan kata yang memiliki konotasi negatif pada setiap komentar yang diambil seperti jel*k, jij*k, bela*gu, tol*l, bang**t, gobl*k, lon*e, dek*l, nor*ak, mony**g baik yang terdapat pada Kamus Besar Bahasa Indonesia (KBBI) ataupun kata-kata yang memang sering dilontarkan sebagai bentuk ejekan atau umpatan serta kalimat yang memang memiliki penyimpangan makna dengan tujuan menghina, merendahkan, menyinggung serta tidak menghormati orang lain baik itu *public figure*, artis maupun selebgram yang dimaksud. Contoh komentar *bullying* dapat dilihat pada tabel 2 sebagai berikut.

TABEL 2
CONTOH KOMENTAR *BULLYING*

Komentar	Penjelasan
"Mukanya kok tua banget ya masih kecil? Itu efek atau bukan kak @lestykejora"	Komentar tersebut seolah bertanya, akan tetapi penyampaian dengan penggunaan kata "tua" yang ditujukan kepada anak kecil terkesan mengejek atau mengolok-olok karena tidak sesuai dengan usia anak dari *public figure* tersebut.

Prosedur Kerja

Berikut merupakan prosedur kerja dari model yang akan dibangun pada penelitian dan dapat dilihat pada gambar 1 sebagai berikut.



Gambar 1. Prosedur Kerja

Adapun tahapan yang akan dilakukan diantaranya yaitu pembacaan *dataset* yang telah dikumpulkan dan disimpan dalam format *.csv*, kemudian akan melalui tahap *Text Preprocessing* untuk membersihkan

dataset tersebut sehingga layak untuk digunakan pada tahap berikutnya. Selanjutnya, yaitu *FastaText Embedding* yang bertujuan untuk mengubah setiap kata yang ada pada data kedalam bentuk *vector*. Kemudian *dataset* tersebut akan dibagi pada tahapan data *splitting*, menjadi data *training* dan data *testing* dengan rasio 80:20. Data yang telah diolah akan melalui tahap klasifikasi menggunakan algoritma *Random Forest Classifier* dan hasilnya akan dievaluasi menggunakan *confusion matrix*. Kemudian dilakukan deteksi komentar sebagai tahap akhir penelitian.

Text Preprocessing

Sebelum diolah dan diklasifikasikan menggunakan model *machine learning* maka akan dilakukan tahap *preprocessing*, tujuannya adalah untuk mempersiapkan data yang akan digunakan agar dapat diproses secara efektif dan efisien oleh model pembelajaran mesin [10]. Dengan melakukan tahap *preprocessing* yang benar maka akan menghasilkan data dengan kualitas yang baik sehingga memungkinkan model pembelajaran mesin menghasilkan klasifikasi dengan prediksi yang lebih akurat dan optimal [11]. Pada penelitian ini data yang digunakan akan melalui empat tahapan *text preprocessing* seperti *cleaning & case folding, tokenization & normalization, stopwords removal* serta *stemming.*

Cleaning bertujuan membersihkan data dengan cara menghilangkan karakter ataupun symbol tertentu yang terdapat pada data untuk mengurangi *noise* [12]. Seperti menghilangkan *URL*, menghapus karakter *non-ASCII*, angka, simbol, tanda baca, *mention*, *hastag*, serta *case folding* yang digunakan untuk mengubah huruf kapital kedalam bentuk *lowercase.* Contoh dari *cleaning & case folding* dapat dilihat pada tabel 3 berikut.

TABEL 3
CLEANING & CASE FOLDING

Sebelum	Sesudah
"Kyknya lebih diubah dibgian bbir aja.. semuanya di @agnezmo msh asli.. liat aja yg slide ahir.. msh natural .. cuma bbir bgian atas yg berbeda"	"kyknya lebih diubah dibgian bbir aja semuanya di agnezmo msh asli liat aja yg slide ahir msh natural cuma bbir bgian atas yg berbeda"
"Yg pda komen "jngn tatoan nez", "mkin ke sni mkin bnyak tatonya", "smnjak sma adam jadi ktgihan tatoan",, duhhh kalo klian emang tau agnes, dari sblum sma adam, agnes dah punya tatto, dh lama lagi, hnya sja di bgian yg jrang klian lihat dan g di publish nya.. di pahanya ada tuh"	"yg pda komen jngn tatoan nez mkin ke sni mkin bnyak tatonya smnjak sma adam jadi ktgihan tatoan duhhh kalo klian emang tau agnes dari sblum sma adam agnes dah punya tatto dh lama lagi hnya sja di bgian yg jrang klian lihat dan g di publishnya di pahanya ada tuh"

Selanjutnya dilakukan proses *tokenization & normalization* yang bertujuan untuk memisahkan setiap kata yang membentuk suatu kalimat menjadi potongan-potongan token [13] serta melakukan normalisasi pada setiap kata berupa singkatan atau *typo* sehingga akan diperbaiki menjadi lebih terstruktur dan dapat diproses lebih lanjut berdasarkan *dictionary* yang telah dipersiapkan sebelumnya. *Dictionary* digunakan untuk memetakan kata-kata yang ingin dinormalisasi ke bentuk standar yang diinginkan, setiap kata dalam teks akan dicocokkan dengan entri dalam *dictionary* dan diubah ke bentuk standar jika ditemukan entri yang sesuai. Contoh *tokenization & normalization* dapat dilihat pada tabel 4 berikut.

TABEL 4
Tokenization & Normalization

Sebelum	Sesudah
"kyknya lebih diubah dibgian bbir aja semuanya di agnezmo msh asli liat aja yg slide ahir msh natural cuma bbir bgian atas yg berbeda"	"sepertinya, lebih, diubah, dibagian, bibir, saja, semuanya, di, agnezmo, masih, asli, lihat, saja, yang, slide, akhir, masih, natural, cuma, bibir, bagian, atas, yang, berbeda"
"yg pda komen jngn tatoan nez mkin ke sni mkin bnyak tatonya smnjak sma adam jadi ktgihan tatoan duhhh kalo klian emang tau agnes dari sblum sma adam agnes dah punya tatto dh lama lagi hnya sja di bgian yg jrang klian lihat dan g di publishnya di pahanya ada tuh"	"yang, pada, komen, jangan, tatoan, nez, semakin, ke, sini, semakin, banyak, tatonya, semenjak, dengan, adam, jadi, ketagihan, tatoan, aduh, kalau, kalian, memang, tahu, agnes, dari, sebelum, dengan, adam, agnes, sudah, punya, tatto, sudah, lama, lagi, hanya, saja, di, bagian, yang, jarang, kalian, lihat, dan, tidak, di, publishnya, di, pahanya, ada, itu"

Jurnal Manajemen Informatika (JAMIKA)
Volume 13 Nomor 1 Edisi April 2023
E ISSN: 2655-6960 | P ISSN: 2088-4125
OJS: https://ojs.unikom.ac.id/index.php/jamika

Tahap *preprocessing* selanjutnya yaitu *stopwords removal* yang bertujuan untuk menghapus kata yang tidak penting dan tidak berpengaruh pada proses klasifikasi [14]. Penelitian ini menggunakan *library NLTK (Natural Language Tool Kit)* pada proses *stopword removal*. Contoh *stopwords removal* dapat dilihat pada tabel 5 berikut.

TABEL 5
Stopwords Removal

Sebelum	Sesudah
"sepertinya, lebih, diubah, dibagian, bibir, saja, semuanya, di, agnezmo, masih, asli, lihat, saja, yang, slide, akhir, masih, natural, cuma, bibir, bagian, atas, yang, berbeda"	"diubah, dibagian ,bibir, agnezmo, asli, lihat, slide, natural, bibir, berbeda"
"yang, pada, komen, jangan, tatoan, nez, semakin, ke, sini, semakin, banyak, tatonya, semenjak, dengan, adam, jadi, ketagihan, tatoan, aduh, kalau, kalian, memang, tahu, agnes, dari, sebelum, dengan, adam, agnes, sudah, punya, tatto, sudah, lama, lagi, hanya, saja, di, bagian, yang, jarang, kalian, lihat, dan, tidak, di, publishnya, di, pahanya, ada, itu"	"komen, tatoan, nez, tatonya, semenjak, adam, ketagihan, tatoan, agnes, adam, agnes, tatto, jarang, lihat, publishnya, pahanya"

Selanjutnya yaitu *stemming* sebagai tahap *preprocessing* terakhir yang bertujuan untuk menemukan kata dasar dengan mengurangi imbuhannya. Imbuhan yang akan dihilangkan seperti imbuhan awalan "me", "ter", "ke", "ber", "di", imbuhan akhiran "kan", "nya", "-i" serta imbuhan lainnya. Penelitian ini menggunakan *library* Sastrawi pada proses *stemming*. Contoh dari *stemming* dapat dilihat pada tabel 6 berikut.

TABEL 6
Stemming

Sebelum	Sesudah
"diubah, dibagian ,bibir, agnezmo, asli, lihat, slide, natural, bibir, berbeda"	"ubah, bagi, bibir, agnezmo, asli, lihat, slide, natural, bibir, beda"
"komen, tatoan, nez, tatonya, semenjak, adam, ketagihan, tatoan, agnes, adam, agnes, tatto, jarang, lihat, publishnya, pahanya"	"komen, tato, nez, tato, semenjak, adam, tagih, tato, agnes, adam, agnes, tatto, jarang, lihat, publish, paha"

Word Embedding FastText

Word embedding adalah kemampuan terdefinisi yang memetakan setiap kata ke dalam vektor berdimensi tinggi. Pengembangan perhitungan pemodelan kata yang didasarkan pada jumlah dan frekuensi kemunculan kata dalam dokumen dikenal sebagai *word embedding*. *Word embedding* mengacu pada kedekatan kontekstual kata atau dokumen berdasarkan data pelatihan yang digunakan dalam pembentukannya; akibatnya, kedekatan ini seringkali tidak mencerminkan makna sebuah kata [15].

FastText adalah salah satu jenis *word embedding* yang merupakan pengembangan dari *word2vec* dengan keunggulan yaitu dapat memperoleh vektor kata yang tidak ada didalam data atau *out of vocabulary* [16]. Pada penelitian ini *dataset* akan diubah kedalam bentuk *sentence vector* dimana nantinya masing-masing komentar akan diwakili oleh vektor dengan panjang dimensi sebesar 100 vektor. Hasil dari *word embedding* dapat dilihat pada gambar 2 berikut.

```
array([[-0.01765488,  0.05324895, -0.20560446, ..., -0.21682529,
         0.15049241, -0.20414221],
       [ 1.3075553 ,  0.11381936,  0.01691299, ..., -0.78896654,
        -0.7541096 , -0.36673516],
       [ 0.5695383 , -0.14695832, -0.05278992, ..., -0.71374536,
         0.431386  , -0.57360595],
       ...,
       [ 0.04031713, -0.24459553,  0.03711206, ...,  0.35999623,
        -0.3761383 ,  0.03401469],
       [ 0.21173215,  0.44759324,  0.31809738, ..., -0.17689788,
        -0.14737165, -0.06766833],
       [ 0.06418078, -0.58173406,  0.10694117, ...,  0.1848037 ,
        -0.59629256, -0.3047408 ]], dtype=float32)
```

Gambar 2. *Sentence Vector Dataset*

Jurnal JAMIKA Manajemen Informatika
Jurnal Manajemen Informatika (JAMIKA)
Volume 13 Nomor 1 Edisi April 2023
E ISSN: 2655-6960 | P ISSN: 2088-4125
OJS: https://ojs.unikom.ac.id/index.php/jamika

Label Encoding

Dalam pemrosesan data, seringka terdapat data dengan nilai-nilai kategori seperti "merah", "biru", atau "hijau" dalam suatu kolom atau atribut. Untuk dapat memproses data ini dengan algoritma pembelajaran mesin, maka perlu mengubah nilai-nilai kategori tersebut menjadi nilai numerik. *Label encoding* adalah salah satu teknik dalam pembelajaran mesin yang digunakan untuk mengubah data kategori menjadi data numerik [17].

Pada *label encoding*, nilai-nilai kategori diubah menjadi bilangan bulat dengan menggunakan suatu aturan tertentu. Misalnya, memberikan label 0 untuk kategori "merah", label 1 untuk kategori "biru", dan label 2 untuk kategori "hijau". Dengan demikian, data kategori sudah dapat diubah menjadi data numerik yang dapat diproses oleh algoritma pembelajaran mesin. Pada penelitian ini *label encoding* digunakan untuk mengubah kelas *bullying* dan *non-bullying* menjadi data numerik yaitu 0 dan 1 agar dapat diproses oleh algoritma *machine learning*. Hasil dari *label encoding* dapat dilihat pada gambar 3 berikut.

	Class	Class_encoded
0	Non-Bullying	1
1	Non-Bullying	1
2	Bullying	0
3	Non-Bullying	1
4	Bullying	0
...	...	...
995	Non-Bullying	1
996	Non-Bullying	1
997	Non-Bullying	1
998	Bullying	0
999	Non-Bullying	1

Gambar 3. *Label Encoding*

Data Splitting

Data splitting adalah teknik dalam pembelajaran mesin yang digunakan untuk membagi *dataset* menjadi dua atau lebih. Pada tahapan data *splitting* data akan dibagi menjadi data *training* dan data *testing*. Dengan jumlah data yang sama pada kedua kelas yaitu 500 kelas *bullying* dan 500 kelas *non-bullying* sehingga total keseluruhan keduanya berjumlah 1000 data. Pembagian data dilakukan dengan menggunakan fungsi *train_test_split*, dimana *dataset* akan dibagi dengan rasio 80:20 sehingga 80% digunakan sebagai data *training* dan 20% digunakan sebagai data *testing*. Rasio 80:20 merupakan pilihan yang umum digunakan karena memberikan hasil yang cukup baik dalam kebanyakan kasus, terutama pada jumlah data yang tidak terlalu besar. Untuk pembagian *dataset* dapat dilihat pada tabel 7 berikut.

TABEL 7
PEMBAGIAN *DATASET*

Kelas	Training	Testing	Jumlah
Bullying	400	100	500
Non-Bullying	400	100	500
Jumlah	800 Komentar	200 Komentar	1000 Komentar

Random Forest Classifier

Algoritma *Random Forest Classifier* merupakan algoritma *ensemble learning*. *Ensemble learning* adalah suatu teknik dengan menggabungkan beberapa model untuk melakukan prediksi [18]. *Random forest classifier* merupakan bentuk implementasi dari *homogeneous ensemble learning* yang menggabungkan beberapa model sejenis yaitu *decision tree* (pohon keputusan) [19]. Dalam algoritma *random forest classifier* dikenal istilah *bootstrap aggregating (bagging)*. Mekanisme *bagging* pada dasarnya menerapkan proses *random sampling with replacement* dimana proses tersebut terletak pada proses *split*. *Split* menghasilkan sejumlah variabel prediktor yang dipilih secara acak sebagai bagian dari proses *split* tersebut [20].

Random forest menyediakan parameter yang dapat disesuaikan dengan data yang akan diklasifikasi. Pada penelitian ini parameter yang digunakan yaitu *random_state*, *n_estimators*, *max_features* dan *criterion*. Nilai parameter akan disesuaikan untuk melakukan beberapa percobaan terhadap model *random forest classifier* yang dibangun.

Jurnal Manajemen Informatika (JAMIKA)
Volume 13 Nomor 1 Edisi April 2023
E ISSN: 2655-6960 | P ISSN: 2088-4125
OJS: https://ojs.unikom.ac.id/index.php/jamika

Berikut cara kerja algoritma *random forest* dalam melakukan klasifikasi:

1. Menggunakan *n* data sampel yang terambil secara acak dari *dataset* yang diberikan dengan menggunakan teknik *bagging* serta pengembalian *(replacement)*. Dalam hal ini *input* berupa *sentence vector* dengan panjang dimensi 100 yang telah dihasilkan dari setiap komentar.

TABEL 8
CONTOH DATA *INPUT*

	X [0]	X [1]	X [2]	……	X [99]
X [0]	-0.017655	0.053249	-0.205604	……	-0.204142
X [1]	1.307555	0.113819	0.016913	……	-0.366735
X [2]	0.569538	-0.146958	-0.052790	……	-0.573606
.	.	.	.	.	.
.	.	.	.	.	.
X[998]	0.211732	0.447593	0.318097	……	-0.067668
X[999]	0.064181	-0.581734	0.106941	……	-0.304741

2. Menggunakan *n* data sampel untuk membangun pohon keputusan, dimana variabel prediktor yang diambil secara acak digunakan untuk memilih node terbaik dalam menentukan split saat membuat pohon keputusan.
3. Melakukan prediksi terhadap sampel *n* berdasarkan pohon yang sudah terbentuk pada tahap 2.
4. Mengulangi langkah 1 sampai langkah 3 hingga *K* kali replikasi.
5. Melakukan *majority voting* untuk menghitung hasil prediksi mayoritas yang dihasilkan dari *K* kali replikasi pada setiap pohon keputusan.
6. Menghitung ketepatan output data *training*.
7. Menghitung ketepatan output data *testing*.
8. Mengulangi langkah 1 sampai dengan langkah 7 dengan mencoba kombinasi jumlah pohon *(K)* yang berbeda yaitu 10, 20, 30 sampai dengan 100.
9. Memilih kombinasi jumlah pohon dengan akurasi paling tinggi.

Evaluasi Model

Evaluasi model pada pembelajaran mesin digunakan untuk mengukur performa model klasifikasi pada dataset yang sudah diketahui labelnya atau dikenal dengan *confusion matrix*. *Confusion matrix* menunjukkan seberapa baik model dapat mengklasifikasikan setiap label yang ada pada *dataset*, dan menunjukkan jumlah prediksi yang benar dan salah [21]. Ilustrasi dari *confusion matrix* dapat dilihat pada gambar 4 sebagai berikut.

		Predicted Class	
		(1) Positive	**(0) Negative**
Actual Class	**(1) Positive**	**TP** (True Positive)	**FN** (False Negative)
	(0) Negative	**FP** (False Positive)	**TN** (True Negative)

Gambar 4. *Confusion Matrix*

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \text{ x } 100\% \quad (1)$$

$$Precission = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1\text{-}Score = \frac{2\ x\ RecallxPrecision}{Recall+Precision} \quad (4)$$

Jurnal Manajemen Informatika (JAMIKA)
Volume 13 Nomor 1 Edisi April 2023
E ISSN: 2655-6960 | P ISSN: 2088-4125
OJS: https://ojs.unikom.ac.id/index.php/jamika

Confusion matrix terdiri dari empat sel atau kotak yaitu *True Positive (TP), False Positive (FP), True Negative (TN)*, dan *False Negative (FN)*. Dari empat sel tersebut dapat dihitung beberapa metrik evaluasi seperti akurasi *(accuracy)*, presisi *(precision)*, *recall (sensitivity)*, dan *f1-score*. Akurasi adalah persentase jumlah prediksi yang benar dari total data, *presisi* adalah persentase data positif yang benar diklasifikasikan, *recall* adalah persentase data positif yang terdeteksi benar oleh model, dan *f1-score* adalah rata-rata harmonik antara *presisi* dan *recall.* Dengan menggunakan *confusion matrix* akan dapat memahami performa model klasifikasi yang dibangun.

III. HASIL DAN PEMBAHASAN

Setelah dilakukan tahap klasifikasi menggunakan algoritma *Random Forest* dengan penyesuaian parameter, maka akan diketahui kombinasi terbaik dari nilai-nilai *hyperparameters* yang digunakan dalam model *Random Forest. Hyperparameters* adalah variabel yang mempengaruhi bagaimana model *Random Forest* dibangun dan beroperasi. Hasil dari *hyperparameters tuning* dapat meningkatkan akurasi dan generalisasi dari model *Random Forest* yang dibangun. Adapun nilai *hyperparameters* yang digunakan pada penelitian ini yaitu parameter *random state* = 0, *n_estimators* = 10-100, *max_features = 'sqrt'*, serta *criterion = 'gini'*.

Dalam mencari *hyperparameters tuning* terbaik digunakan *trial and error method* yaitu dengan melakukan beberapa kali percobaan untuk mendapatkan hasil yang diinginkan sehingga dapat diketahui kombinasi parameter terbaik untuk dihitung nilai *precision, recall, f1_score* dan akurasinya. Performansi *hyperparameteres tuning* dapat dilihat pada tabel 9 dimana B merepresentasikan *Bullying* dan NB merepresentasikan *Non-Bullying*.

TABEL 9
PERFORMA *HYPERPARAMETERES TUNING*

random state	n_estimators	max_ features	criterion	Precision		Recall		F1_Score		Accuracy
				B	NB	B	NB	B	NB	
0	10	'sqrt'	'gini'	71	78	81	67	76	72	74%
0	20	'sqrt'	'gini'	74	83	86	69	79	75	78%
0	30	'sqrt'	'gini'	77	85	87	74	82	79	81%
0	40	'sqrt'	'gini'	77	88	90	73	83	80	81%
0	50	'sqrt'	'gini'	78	85	87	75	82	80	81%
0	60	'sqrt'	'gini'	79	84	85	77	82	80	81%
0	70	'sqrt'	'gini'	80	82	83	79	81	81	81%
0	80	'sqrt'	'gini'	80	84	85	79	83	81	82%
0	90	'sqrt'	'gini'	80	86	87	78	83	82	82%
0	100	'sqrt'	'gini'	84	85	85	84	85	84	84%

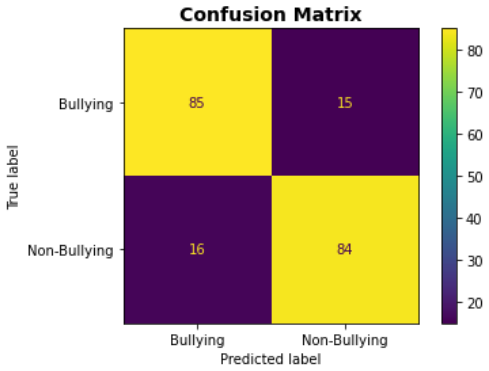
Berdasarkan tabel diatas dapat dilihat performa *hyperparameteres tuning* pada algoritma *Random Forest* dalam melakukan klasifikasi sehingga diperoleh hasil akurasi terbaik yang terdapat pada kombinasi *hyperparameteres tuning* terakhir dimana *n_estimator* = 100, *max_features* = 'sqrt' dan *criterion* = 'gini' dengan nilai akurasi sebesar 84%. Berikut merupakan hasil pengujian dari data *testing* pada kombinasi *hyperparameteres tuning* terakhir yang dapat dilihat pada tabel 10.

TABEL 10
HASIL PENGUJIAN DATA *TESTING*

No	Comment	Class	Pred
1	kakak ini orgnya keliatan ramahnya bgt dan jujur	Non-Bullying	Non-Bullying
2	aduh cantik ny kebangetan sih ini heehe	Non-Bullying	Non-Bullying
3	Menurutku pribadi, dia tulus minta maafnya….	Non-Bullying	Non-Bullying
4	kasian billar anaknya mirip bgt lesti wkwk idung nya pesek gt njir	Bullying	Bullying
5	The best banget produknya Batrisyia bener bener nyata khasiat dan manfaatnya	Non-Bullying	Non-Bullying
6	Kek Pegang Boneka Annabelle Versi Lite	Bullying	Bullying
7	manusia gak ada moral, jahat, si tukang sogok	Bullying	Bullying
8	Si monyong bibir tempe	Bullying	Bullying
...	…	…	…

Jurnal Manajemen Informatika (JAMIKA)
Volume 13 Nomor 1 Edisi April 2023
E ISSN: 2655-6960 | P ISSN: 2088-4125
OJS: https://ojs.unikom.ac.id/index.php/jamika

No	Comment	Class	Pred
194	netizen tuh ngurusin bgt hidup org, mau dia unfollow siapa aja bukan urusan kalian, jgn trllu jauh guys ngatur hidup orang	Non-Bullying	Bullying
195	kaaamuuu gaaatal gatal gatal bukannya digaruk malah…..	Bullying	Bullying
196	mungkin ini ujian omay .. semua kejadian ttp ada hikmah dan pelajaran ya okay ttp jadi orang baik	Non-Bullying	Non-Bullying
197	Jangan di bales komen hate bang bikin sakit jempol, mending lanjut berkarya:)	Non-Bullying	Non-Bullying
198	Dibalik ketawa nya. Terlihat muka sedih...tetap semangat marshel dalam mencari rezeki yg halal	Non-Bullying	Non-Bullying
199	MATI KAU NAJIS KONTOL BANGSAT MATI KAU	Bullying	Bullying
200	Jempolnya knpa harus gitu sih dek nangisss Gemes bgt loh	Non-Bullying	Bullying



Gambar 5. *Confusion Matrix Data Testing*

Pada gambar 5 menampilkan visualisasi hasil klasifikasi dalam bentuk *heatmap* untuk mempermudah dalam melihat pola TP (*True Positif*), TN (*True Negatif*), FP (*False Positif*), FN (*False Negatif*) dari data *testing* dimana terdapat 85 komentar *bullying* yang diprediksi sebagai *bullying*, 15 komentar *bullying* yang diprediksi sebagai *non-bullying*, 84 komentar *non-bullying* yang diprediksi sebagai *non-bullying* dan 16 komentar *non-bullying* yang diprediksi sebagai *bullying*, sehingga dapat diambil kesimpulan berupa pola *TP*, *TN*, *FP*, *FN* yang dapat dilihat pada tabel 11 sebagai berikut.

TABEL 11
TP, *TN*, *FP*, *FN* DATA *TESTING*

	Prediksi Komentar Bullying (0)	Prediksi Komentar Non-Bullying (1)
TP	85	84
TN	84	85
FP	16	15
FN	15	16

Selanjutnya dilakukan perhitungan secara manual untuk memperoleh nilai *precision*, *recall*, *f1_score* dan *accuracy* berdasarkan tabel 11 menggunakan persamaan sebagai berikut.

1. *Precision Bullying* dan *Non-Bullying*

   Precision komentar *Bullying* (0) = $\frac{TP0}{TP0+FP0} = \frac{85}{85+16} = 0.84$ (5)

   Precision komentar *Non-Bullying* (1) = $\frac{TP1}{TP1+FP1} = \frac{84}{84+15} = 0.85$ (6)

2. *Recall Bullying* dan *Non-Bullying*

   Recall komentar *Bullying* (0) = $\frac{TP0}{TP0+FN0} = \frac{85}{85+15} = 0.85$ (7)

Jurnal Manajemen Informatika (JAMIKA)
Volume 13 Nomor 1 Edisi April 2023
E ISSN: 2655-6960 | P ISSN: 2088-4125
OJS: https://ojs.unikom.ac.id/index.php/jamika

Recall komentar *Non-Bullying* (1) = $\frac{TP1}{TP1+FN1} = \frac{84}{84+16} = 0.84$ (8)

3. *F1_Score Bullying* dan *Non-Bullying*
   F1_Score komentar *Bullying* (0) = $\frac{2\ x\ Recall(0) x Precision(0)}{Recall(0)+Precision(0)} = \frac{2\ x\ 0{,}85\ x\ 0{,}84}{0{,}85+0{,}84} = 0.84$ (9)

   F1_Score komentar *Non-Bullying* (1) = $\frac{2\ x\ Recall(1) x Precision(1)}{Recall(1)+Precision(1)} = \frac{2\ x\ 0{,}84\ x\ 0{,}85}{0{,}84+0{,}85} = 0.84$ (10)

4. Total Akurasi Keseluruhan
   Accuracy = $\frac{TP+TN}{TP+TN+FP+FN} = \frac{169}{200}$ x 100% = 84% (11)

Berdasarkan hasil perhitungan manual nilai *precision, recall, f1_score* dan *accuracy* terhadap komentar *bullying* dan *non-bullying* diatas dapat dilihat bahwa nilai *precision* pada komentar *non-bullying* lebih tinggi, yaitu 85% dan nilai *precision* pada komentar *bullying* lebih rendah yaitu 84%. Sebaliknya nilai *recall* pada komentar *bullying* lebih tinggi yaitu 85% dan nilai *recall* pada komentar *non-bullying* lebih rendah yaitu 84%. Sedangkan untuk nilai *f1_score* pada komentar *bullying* dan *non_bullying* keduanya memperoleh nilai yang sama, yaitu 84% dengan menghasilkan nilai akurasi yang cukup baik sebesar 84%.

Selanjutnya dilakukan pengecekan terhadap komentar baru yang akan dideteksi apakah komentar tersebut termasuk kedalam *class bullying* atau *non-bullying* sebagai tahap akhir penelitian yang dapat dilihat pada gambar 6 sebagai berikut.

```
1 Deteksi("uda tua kelakuan kayak bocah")
'Bullying'

1 Deteksi("kakak baik banget, cantik lagi")
'Non-Bullying'
```

Gambar 6. Deteksi Komentar *Bullying* dan *Non-Bullying*

Model yang dibangun sudah dapat melakukan prediksi komentar *bullying* dan *non-bullying* dengan baik, akan tetapi kekurangan pada penelitian ini, yaitu masih terdapat komentar yang diprediksi tidak sesuai dengan semestinya, nilai akurasi yang diperoleh juga mempengaruhi keakuratan saat melakukan prediksi terhadap komentar baru.

IV. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan dan dipaparkan pada hasil dan pembahasan mengenai deteksi komentar *cyberbullying* pada media sosial *instagram* menggunakan algoritma *Random Forest* dapat ditarik kesimpulan, yaitu nilai akurasi terbaik dari algoritma *Random Forest* terdapat pada kombinasi *hyperparameteres tuning* terakhir dengan parameter *n_estimator* = 100 dimana akurasi yang diperoleh sebesar 84% dan telah dihitung berdasarkan hasil evaluasi *confusion matrix.* Selain itu model yang dibangun juga dapat mendeteksi komentar baru dengan hasil prediksi yang cukup baik walaupun masih terdapat komentar yang diprediksi tidak sesuai dengan semestinya. Adapun saran yang dapat diterapkan untuk mengembangkan penelitian selanjutnya, yaitu melakukan klasifikasi komentar *cyberbullying* dengan kelas yang lebih spesifik seperti komentar *rasisme* atau komentar *seksisme* serta mempertimbangkan penggunaan *embedding* kata yang lain untuk mendapatkan hasil yang lebih baik.

DAFTAR PUSTAKA

[1] APJII, "Laporan Survei Internet APJII 2019 – 2020," *Asos. Penyelenggara Jasa Internet Indones.*, vol. 2020, pp. 1–146, 2020, [Online]. Available: https://apjii.or.id/survei.

[2] A. Perwira, J. Dwitama, and K. Kunci, "Deteksi Ujaran Kebencian Pada Twitter Bahasa Indonesia Menggunakan Machine Learning : Reviu Literatur," *J. SNATi*, vol. 1, no. 1, pp. 31–39, 2021.

[3] R. Yunanto, A. P. Purfini, and A. Prabuwisesa, "Survei Literatur: Deteksi Berita Palsu Menggunakan Pendekatan Deep Learning," *J. Manaj. Inform.*, vol. 11, no. 2, pp. 118–130, 2021, doi:

10.34010/jamika.v11i2.5362.

[4] A. S. Hutagalung, A. B. P. Negara, and E. E. Pratama, "Aplikasi Pendeteksi Cyberbullying Terhadap Komentar Postingan Media Sosial Instagram dengan Metode Naïve Bayes Classifier Berbasis Website," *J. Sist. dan Teknol. Inf.*, vol. 9, no. 3, p. 364, 2021, doi: 10.26418/justin.v9i3.44843.

[5] J. Pardede, Y. Miftahuddin, and W. Kahar, "Deteksi Komentar Cyberbullying Pada Media Sosial Berbahasa Inggris Menggunakan Naïve Bayes Classification," *J. Inform.*, vol. 7, no. 1, 2020, [Online]. Available: http://ejournal.bsi.ac.id/ejurnal/index.php/ji.

[6] H. Junawan and N. Laugu, "Eksistensi Media Sosial,Youtube, Instagram dan Whatsapp Ditengah Pandemi Covid-19 Dikalangan Masyarakat Virtual Indonesia," *Baitul 'Ulum J. Ilmu Perpust. dan Inf.*, vol. 4, no. 1, pp. 41–57, 2020, doi: 10.30631/baitululum.v4i1.46.

[7] D. Riswanto and R. Marsinun, "Perilaku Cyberbullying Remaja di Media Sosial," *Analitika*, vol. 12, no. 2, pp. 98–111, 2020, doi: 10.31289/analitika.v12i2.3704.

[8] D. I. Cahyani, M. Politeknik, N. Lhokseumawe, and A. Pendahuluan, "Cyberbullying Di Media Sosial Dalam Perspektif Al- Qur ' an," *Ilmu Al-Qur'an dan Tafsir*, vol. 1, no. 1, pp. 36–51, 2022.

[9] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 2, p. 640, Apr. 2021, doi: 10.30865/mib.v5i2.2937.

[10] R. M. Candra and A. Nanda Rozana, "Klasifikasi Komentar Bullying pada Instagram Menggunakan Metode K-Nearest Neighbor," *IT J. Res. Dev.*, vol. 5, no. 1, pp. 45–52, Jul. 2020, doi: 10.25299/itjrd.2020.vol5(1).4962.

[11] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.

[12] S. Khomsah and Agus Sasmito Aribowo, "Model Text-Preprocessing Komentar Youtube Dalam Bahasa Indonesia," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 4, pp. 648–654, 2020.

[13] T. Nugraha Manoppo and D. Hatta Fudholi, "Deteksi Cyberbullying berdasarkan Unsur Perbuatan Pidana yang Dilanggar dengan Naive Bayes dan Support Vector Machine," 2021.

[14] R. Riyaddulloh and A. Romadhony, "Normalisasi Teks Bahasa Indonesia Berbasis Kamus Slang Studi Kasus: Tweet Produk Gadget Pada Twitter," *eProceedings Eng.*, vol. 8, no. 4, pp. 4216–4228, 2021,[Online].Available:https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15246/14969.

[15] H. Utama and A. Masruro, "Analisis Sentimen pada Twitter Menggunakan Word Embedding dengan Pendekatan Word2Vec," *J. Sist. Cerdas*, vol. 05, no. 02, pp. 128–134, 2022.

[16] A. Nurdin, B. Anggo Seno Aji, A. Bustamin, and Z. Abidin, "Perbandingan Kinerja Word Embedding Word2Vec, Glove, Dan Fasttext Pada Klasifikasi Teks," *J. Tekno Kompak*, vol. 14, no. 2, p. 74, 2020, doi: 10.33365/jtk.v14i2.732.

[17] P. Purwono, A. Wirasto, and K. Nisa, "Comparison of Machine Learning Algorithms for Classification of Drug Groups," *Sisfotenika*, vol. 11, no. 2, p. 196, 2021, doi: 10.30700/jst.v11i2.1134.

[18] A. U. Zailani and N. L. Hanun, "Penerapan Algoritma Klasifikasi Random Forest Untuk Penentuan Kelayakan Pemberian Kredit Di Koperasi Mitra Sejahtera," *Infotech J. Technol. Inf.*, vol. 6, no. 1, pp. 7–14, 2020, doi: 10.37365/it.v6i1.61.

[19] S. Budiman, A. Sunyoto, and A. Nasiri, "Analisa Performa Penggunaan Feature Selection untuk Mendeteksi Intrusion Detection Systems dengan Algoritma Random Forest Classifier," *Sistemasi*, vol. 10, no. 3, p. 753, 2021, doi: 10.32520/stmsi.v10i3.1550.

[20] I. K. P. Suniantara, "ANALISIS RANDOM FOREST PADA KLASIFIKASI CART UNIVERSITAS TERBUKA Analysis of Random Forest In Inaccuracies CART Classification of Terbuka University Student Graduates," vol. 13, no. 3, pp. 179–186, 2019.

[21] Z. Firmansyah and N. F. Puspitasari, "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Berdasarkan Opini Pada Twitter Menggunakan Algoritma Naive Bayes," *J. Tek. Inform.*, vol. 14, no. 2, pp. 171–178, 2021, [Online]. Available: https://doi.org/10.15408/jti.v14i2.24024.