

Comparison of Algorithms for Cyberbullying Detection to Football Player in Social Media

Perbandingan Algoritma untuk Deteksi Cyberbullying pada Pemain Sepakbola di Media Sosial

Pawit Widiyantoro¹, Rosita Dian Febriyanti²

^{1,2} Informatika, Institut Teknologi Telkom Purwokerto, Indonesia

¹21102184@ittelkom-pwt.ac.id, ^{2*}21102186@ittelkom-pwt.ac.id

*: Penulis korespondensi (corresponding author)

Informasi Artikel

Received: December 2023
Revised: June 2024
Accepted: September 2024
Published: October 2024

Abstract

Purpose: to compare Naïve Bayes, Support Vector Machine (SVM), and K-Nearest Neighbor (KNN) algorithms for detecting cyberbullying that happen to football player in social media.

Design/methodology/approach: In the cyberbullying detection process, the steps involved are data collection, labeling, data preprocessing, feature extraction, modeling, and finally evaluation by comparing the accuracy values of the three methods used.

Findings/result: Based on the accuracy values obtained, Naive Bayes emerged as the algorithm with the highest accuracy at 78.6%, followed by Support Vector Machine (SVM) with an accuracy of 77.9%, and K-Nearest Neighbor (KNN) with an accuracy of 65.6%.

Originality/value/state of the art: This research discusses the comparison of algorithms for detecting cyberbullying in social media related to football players, an area that has not been addressed by other studies. Additionally, the preprocessing stage and the three algorithms used were also designed and chosen by the researchers themselves.

Abstrak

Tujuan: membandingkan algoritma Naïve Bayes, Support Vector Machine (SVM), dan K-Nearest Neighbor (KNN) dalam mendeteksi cyberbullying pada pemain sepakbola di media sosial.

Keywords: Natural Language Processing (NLP); Cyberbullying Detection; Machine Learning
Kata kunci: Natural Language Processing (NLP); Deteksi Cyberbullying; Machine Learning

Perancangan/metode/pendekatan: pada proses deteksi *cyberbullying*, tahapan yang dilakukan adalah pengumpulan data, *labeling*, *pre-processing* data, *feature extraction*, *modeling*, dan terakhir evaluasi dengan membandingkan nilai akurasi dari ketiga metode yang digunakan.

Hasil: Berdasarkan nilai akurasi yang didapatkan, *Naïve Bayes* menjadi algoritma dengan akurasi tertinggi dengan akurasi 78,6%, diikuti *Support Vector Machine (SVM)* dengan akurasi 77,9%, dan *K-Nearest Neighbor (KNN)* dengan akurasi 65,6%.

Keaslian/ *state of the art*: Penelitian ini membahas tentang perbandingan algoritma untuk deteksi *cyberbullying* di media sosial terhadap pemain sepakbola, dimana belum ada penelitian lain yang membahasnya. Sebagai tambahan, tahap *pre-processing* dan tiga algoritma yang digunakan juga dirancang dan dipilih oleh peneliti sendiri.

1. Pendahuluan

Cyberbullying merupakan sebuah aksi untuk menakuti, menyerang, atau membully seseorang melalui cara berkomunikasi modern dengan media sosial [1]. *Cyberbullying* dianggap sebagai kasus *cybercrime* yang serius karena dapat menyebabkan kekerasan emosional terhadap korban [2]. *Cyberbullying* tidak hanya sebatas pada membuat identitas palsu dan memposting foto atau video yang memalukan atau rumor yang tidak mengenakan tentang seseorang, memberi ancaman juga termasuk ke dalam *cyberbullying* [1]. *Cyberbullying* dapat berujung pada aktivitas menyakiti diri sendiri dan bunuh diri [2]. Dalam kasus yang berat, *bullying* dapat menjadi pemicu tindakan bunuh diri dan tindakan fatal lainnya [3].

Media sosial memungkinkan terjadinya *cyberbullying* yang menyebabkan dampak fisik maupun mental pada korbannya [4]. Beberapa media sosial yang paling sering menjadi tempat terjadinya *cyberbullying* diantaranya adalah Twitter, Instagram, Facebook, Youtube, Snapchat, Skype, dan Wikipedia [5]. Sekitar 21% remaja mengatakan alasan kenapa mereka selalu mengecek media sosial adalah untuk memastikan tidak ada yang mengatakan hal buruk tentang mereka [6]. Seringkali para pelaku menggunakan nama palsu dan tidak memiliki hubungan langsung dengan korban [7].

Cyberbullying dalam sepakbola dan olahraga lainnya bukanlah merupakan fenomena baru [8]. Fans sepak bola dikenal sering mengekspresikan kekecewaan mereka terhadap pemain tertentu ketika klub yang mereka dukung kalah melalui media sosial [9]. Sebagai contoh, pada Mei 2021 Marcus Rashford menerima kurang lebih 70 pesan *abusive* di media sosial setelah kekalahan Manchester United di final *Europa League* yang berisi emoji monyet [8]. Terjadi kegagalan yang sistematis di dalam sepakbola dalam merespon *bullying* terutama rasisme di sosial media. Kegagalan ini terdiri atas, kurangnya koordinasi tentang tanggung jawab dari institusi, organisasi, dan klub [10].

Oleh karena itu, jelas bahwa ketersediaan sistem yang secara otomatis dapat mengidentifikasi perilaku yang diklasifikasikan sebagai *cyberbullying* dapat sangat berguna untuk mencegah

“resiko” dari korbannya [7]. Sistem deteksi *cyberbullying* akan sangat berguna untuk media sosial dan aplikasi *messaging* lainnya untuk melacak konten *cyberbullying* dan mengurangi kasus *cyberbullying*. Tujuan utama dari sistem pendeteksi *cyberbullying* adalah untuk mengidentifikasi teks *cyberbullying* dengan mempertimbangkan arti kalimat tersebut [5]. Deteksi kejadian *cyberbullying* yang dicapai dengan menganalisis konten media sosial merupakan teknik dasar yang digunakan untuk mencegah kejadian *cyberbullying* [2]. Metode pendeteksian harus mampu mengidentifikasi adanya *cyberbullying* dan jenis jenisnya dari berbagai aktivitas di media social [4].

Implementasi *Machine Learning* dalam bidang deteksi *cyberbullying* dapat dilakukan dengan berbagai Algoritma seperti Naïve Bayes, SVM, KNN, atau bahkan metode lainnya seperti Neural Networks dan Transformers [11]. Tahap *preprocessing* yang dilakukan secara umum dalam pemrosesan data teks, yaitu dimulai dari proses *case folding*, *tokenizing*, *cleaning*, *penghapusan stopwords*, dan *stemming* [12]. Penelitian ini bertujuan untuk membandingkan akurasi dari tiga algoritma *Machine Learning* yaitu Naïve Bayes, Support Vector Machine (SVM), dan K-Nearest Neighbor (KNN) dalam mengklasifikasi *cyberbullying* kepada pemain sepakbola di media sosial.

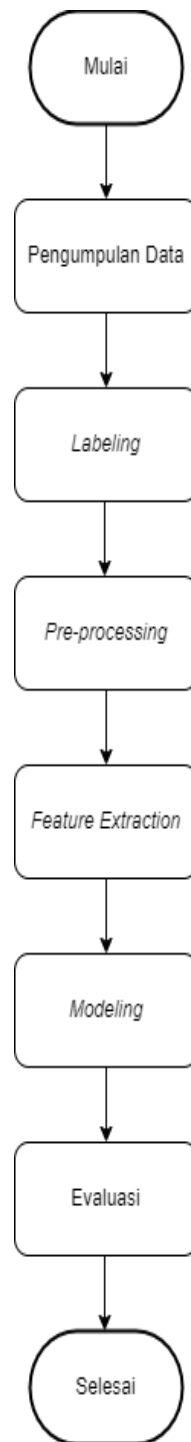
2. Metode/Perancangan

Pada metode dibahas tentang metode, tahapan, maupun model yang digunakan dalam penelitian yang dilakukan. Berikut adalah gambaran alur proses (*flowchart*) dari metode perancangan pada **Gambar 1**.

2.1. Pengumpulan Data

Langkah pertama dalam penelitian ini adalah pengumpulan data dari platform media sosial, dalam hal ini dari postingan dan komentar dari platform media sosial Twitter (X) [13]. Data dikumpulkan dengan menggunakan *Tool Harvest*. Data berisi *tweet* dan balasan yang membahas tentang para pemain sepak bola baik lokal maupun internasional. Digunakan keyword para pemain yang sering terlibat kontroversi yang memungkinkan terjadinya aktivitas *cyberbullying* sebagai sumber dari kelas *cyberbullying* seperti Onana, Arhan, Dalot, Nadeo, dll. Sementara untuk kelas bukan *cyberbullying* digunakan keyword pemain yang sering menuai pujian publik seperti Haaland, De gea, Courtois, dll.

Data diatur sedemikian rupa agar berasal dari wilayah Indonesia menggunakan Geocode wilayah Indonesia. Selain dari pencarian acak *tweet* tentang topik tertentu, dilakukan pula cara lain dengan mengumpulkan data dari *thread* yang memiliki banyak balasan sesuai topik yang dibahas. Didapatkan total data sebanyak 5490 data. **Tabel 1** dibawah ini merupakan contoh data yang telah dikumpulkan.



Gambar 1. Flowchart Metode/Perancangan

Tabel 1. Contoh Data

created_at	id_str	full_text	quote_count	reply_count	retweet	favorite	lang	user_id_str	username	tweet_url
Mon Dec 04 09:10:58 +00 2023	1731601982883090000	ðŸŸš BREAKING : Pemain Timnas Indonesia Pratama Arhan dipastikan bakal meninggalkan Tokyo Verdy. Kepastian itu disampaikan oleh sang agen Dusan Bogdanovic. Kontraktannya sendiri disebut akan berakhir pada akhir tahun 2023. Arhan kabarnya akan pindah ke klub kastaâ€ https://t.co/oexkOLIQa5	1	7	2	5	id	1222927522033090000	idextratime	https://twitter.com/idextratime/status/1731601982883090556
Sun Dec 10 14:21:35 +00	1733854476275200000	Arhan harusnya gandeng kiri kanan sik kasian asnawi sendirian	2	1	3	4	id	1295378909467310000	riris12via	https://twitter.com/riris12via/status/1733854476275204474

created_at	id_str	full_text	quote_count	reply_count	retweet	favorite	lang	user_id_str	username	tweet_url
00 202 3		wkwkwkw kww https://t.co/ KQ859SH Evs								
Mo n Dec 11 15: 10: 20 +00 00 202 3	17342291 33255900 000	Satu resto disewa kayanya buat Arhan kumpul sama temen2 nya ðŸŒŒ mana yang jadi tukang foto mereka zize lagi wkwkwk https://t.co/ 39KHILrnI 1	0	2	3	1 0 7	i n	73109523 14923290 00	mostgirl _aya	https://twitter.com/mostgirl_aya/status/1734229133255905540
Mo n Dec 11 11: 52: 30 +00 00 202 3	17341793 45554690 000	Aku nihh dul @akuabdul dul Mending di surut maen jadi bek ky hariono sih sleding sana sini biar g tembus Dari pada harus bucin kaya arhan gini apalagi pke baju emyu duhh gak dlu Fans emyu nya kaya kmu aja nt mulu arhan doang yg sukses	1	0	0	0	i n	11977730 43587670 000	muham madbra gim	https://twitter.com/muhammadbragim/status/1734179345554698481

created_at	id_str	full_text	quote_count	reply_count	retweet	favorite	lang	user_id_str	username	tweet_url
		bucinnyað Ÿ~- https://t.co/ MvAfKJm 4R5								

2.2. Labeling

Digunakan *tool Gensim* untuk melakukan *labeling* secara otomatis pada data yang telah dikumpulkan. *Labeling* manual juga dilakukan untuk memperkuat hasil yang telah didapatkan dari teknik *labeling* otomatis.

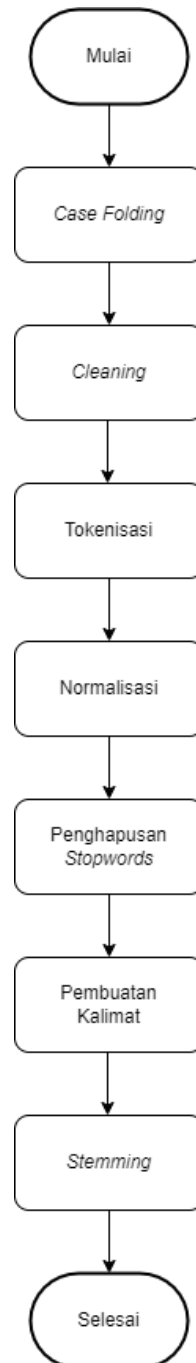
Hasil labeling menunjukkan dari total 5490 data terdapat data dengan kelas *non-cyberbullying* sebanyak 2803 data dan kelas *cyberbullying* sebanyak 2687 data. **Tabel 2** dibawah ini merupakan contoh hasil *labeling* data yang telah dikumpulkan dan diberi label 0 (*non-cyberbullying*) dan 1 (*cyberbullying*).

Tabel 2. Contoh Hasil *Labeling* Data

created_at	Dominan_ Topic	full_text	username
Mon Dec 04 09:10:58 +0000 2023	0	ðŸš™ BREAKING : Pemain Timnas Indonesia Pratama Arhan dipastikan bakal meninggalkan Tokyo Verdy. Kepastian itu disampaikan oleh sang agen Dusan Bogdanovic. Kontrak Pratama Arhan sendiri disebut akan berakhir pada akhir tahun 2023. Arhan kabarnya akan henggang ke klub kastaâ€¦ https://t.co/oexkOLIQa5	idextratime
Sun Dec 10 14:21:35 +0000 2023	0	Arhan harusnya gandeng kiri kanan sik kasian asnawi sendirian wkwkwkwkwk https://t.co/KQ859SHEvs	riris12via
Mon Dec 11 15:10:20 +0000 2023	0	Satu resto disewa kayanya buat Arhan kumpul sama temen2 nya ðŸŒŒ mana yang jadi tukang foto mereka zize lagi wkwkwk https://t.co/39KHILmI1	mostgirl_aya
Mon Dec 11 11:52:30 +0000 2023	1	Aku nihh dul @akuabduldul Mending di surut maen jadi bek ky hariono sih sleding sana sini biar g tembus Dari pada harus bucin kaya arhan gini apalagi pke baju emyu duhh gak dlu Fans emyu nya kaya kamu aja nt mulu arhan doang yg sukses bucinnyaðŸ~- https://t.co/MvAfKJm4R5	muhammadbra gim

2.3. Pre-processing

Berikut merupakan alur/flowchart dari *pre-processing* pada **Gambar 2**.



Gambar 2. Flowchart Pre-processing

Data yang telah melalui proses *labeling* kemudian akan menjalani proses *pre-processing*. Proses *pre-processing* dilakukan dengan mengimplementasikan beberapa teknik *Natural Language*

Processing (NLP) seperti, *case folding*, *cleaning*, tokenisasi, penghapusan *stopwords*, dan *stemming*.

2.3.1. Case Folding

Case folding merupakan proses menyeragamkan semua huruf dari data menjadi huruf kecil [3]. Berikut merupakan contoh dari *case folding* pada **Tabel 3**.

Tabel 3. *Case Folding*

full_text	case_folding
ðŸš” BREAKING : Pemain Timnas Indonesia Pratama Arhan dipastikan bakal meninggalkan Tokyo Verdy. Kepastian itu disampaikan oleh sang agen Dusan Bogdanovic. Kontrak Pratama Arhan sendiri disebut akan berakhir pada akhir tahun 2023. Arhan kabarnya akan henggang ke klub kastaâ€¦ https://t.co/oexkOLIQa5	ðŸš” breaking : pemain timnas indonesia pratama arhan dipastikan bakal meninggalkan tokyo verdy. kepastian itu disampaikan oleh sang agen dusan bogdanovic. kontrak pratama arhan sendiri disebut akan berakhir pada akhir tahun 2023. arhan kabarnya akan henggang ke klub kastaâ€¦ https://t.co/oexkOLIQa5

2.3.2. Tokenisasi

Tokenisasi merupakan proses merubah kalimat menjadi kata-kata tunggal atau token [3]. Proses tokenisasi dilakukan menggunakan kelas *word_tokenize()* yang berasal dari *library NLTK* (*Natural Language Tool Kit*). Berikut contoh hasil tokenisasi pada **Tabel 4**.

Tabel 4. Tokenisasi

full_text	Tokenizing
ðŸš” BREAKING : Pemain Timnas Indonesia Pratama Arhan dipastikan bakal meninggalkan Tokyo Verdy. Kepastian itu disampaikan oleh sang agen Dusan Bogdanovic. Kontrak Pratama Arhan sendiri disebut akan berakhir pada akhir tahun 2023. Arhan kabarnya akan henggang ke klub kastaâ€¦ https://t.co/oexkOLIQa5	['breaking', 'pemain', 'timnas', 'indonesia', 'pratama', 'arhan', 'dipastikan', 'bakal', 'meninggalkan', 'tokyo', 'verdy', 'kepastian', 'itu', 'disampaikan', 'oleh', 'sang', 'agen', 'dusan', 'bogdanovic', 'kontrak', 'pratama', 'arhan', 'sendiri', 'disebut', 'akan', 'berakhir', 'pada', 'akhir', 'tahun', 'arhan', 'kabarnya', 'akan', 'henggang', 'ke', 'klub', 'kasta']

2.3.3. Cleaning

Proses *cleaning* bertujuan membersihkan data raw dari komponen yang tidak berhubungan dengan informasi yang ingin diambil [3]. Data yang didapat memiliki teks yang *noisy*, sehingga perlu dibersihkan terlebih dahulu. Beberapa tindakan yang dilakukan terhadap data diantaranya:

- Menghapus *hashtag* (#)
- Menghapus *mentions* (@)
- Menghapus *retweet* (RT)
- Menghapus link (<https://...>)
- Menghapus karakter yang bukan termasuk alfabet
- Mengganti baris baru dengan spasi
- Menghapus semua tanda baca

- Menghapus spasi di sebelah kanan dan kiri teks

Semua tindakan tadi dilakukan dengan memanfaatkan *regex* dan dikelompokkan menjadi satu fungsi. Berikut contoh hasil dari proses cleaning pada **Tabel 5**.

Tabel 5. *Cleaning*

full_text	cleaning
ðŸš” BREAKING : Pemain Timnas Indonesia Pratama Arhan dipastikan bakal meninggalkan Tokyo Verdy. Kepastian itu disampaikan oleh sang agen Dusan Bogdanovic. Kontrak Pratama Arhan sendiri disebut akan berakhir pada akhir tahun 2023. Arhan kabarnya akan henggang ke klub kastaâ€ https://t.co/oexkOLIQa5	breaking pemain timnas indonesia pratama arhan dipastikan bakal meninggalkan tokyo verdy kepastian itu disampaikan oleh sang agen dusan bogdanovic kontrak pratama arhan sendiri disebut akan berakhir pada akhir tahun arhan kabarnya akan henggang ke klub kasta

2.3.4. Normalisasi

Normalisasi atau kerap dipanggil juga sebagai *filtering* merupakan proses merubah kata dalam bentuk tidak baku atau kata *slang* ke dalam kata baku sesuai Kamus Besar Bahasa Indonesia (KBBI) [3]. Beberapa kata *slang* dalam Bahasa Indonesia diubah ke dalam bentuk bakunya. Kata *slang* yang dimaksud dimasukkan ke dalam sebuah file json agar mampu menampung banyak kata sekaligus. Namun, meskipun sudah banyak kata yang dimasukkan ke dalam list kata *slang*, normalisasi teks masih belum berjalan maksimal. Hal tersebut dikarenakan variasi penggunaan kata *slang* yang dimaksud, seperti penggunaan huruf yang lebih banyak dan variasi bentuk kata. Sebagai contoh terdapat kata “anjingg” dan “anjinggg” yang mana memiliki arti yang sama dan juga kata “wkwkwk” dan “wkwkkw” yang juga memiliki arti sama tetapi berbeda dalam variasi penulisannya.

Berikut contoh hasil normalisasi pada **Tabel 6**.

Tabel 6. Normalisasi

full_text	normalisasi_kata
ðŸš” BREAKING : Pemain Timnas Indonesia Pratama Arhan dipastikan bakal meninggalkan Tokyo Verdy. Kepastian itu disampaikan oleh sang agen Dusan Bogdanovic. Kontrak Pratama Arhan sendiri disebut akan berakhir pada akhir tahun 2023. Arhan kabarnya akan henggang ke klub kastaâ€ https://t.co/oexkOLIQa5	['breaking', 'pemain', 'timnas', 'indonesia', 'pratama', 'arhan', 'dipastikan', 'bakal', 'meninggalkan', 'tokyo', 'verdy', 'kepastian', 'itu', 'disampaikan', 'oleh', 'sang', 'agen', 'dusan', 'bogdanovic', 'kontrak', 'pratama', 'arhan', 'sendiri', 'disebut', 'akan', 'berakhir', 'pada', 'akhir', 'tahun', 'arhan', 'kabarnya', 'akan', 'henggang', 'ke', 'klub', 'kasta']

2.3.5. Penghapusan Stopwords

Stopwords merupakan kumpulan kata yang tidak memiliki arti penting dalam kalimat, *stopwords* harus dihilangkan agar akurasi yang dihasilkan menjadi lebih baik. Sama seperti kata *slang*, *stopwords* juga dibuatkan sebuah file sendiri agar mampu menampung lebih banyak kata.

Berikut beberapa contoh *stopword* yang dihilangkan. Sebagai contoh yaitu kata “ada”, “adalah”, “adanya”, “adapun”, “agak”, “agakny”, dan “agar”. Berikut contoh hasil penghapusan *stopwords* pada **Tabel 7**.

Tabel 7. Penghapusan *Stopwords*

full_text	stopword
ðŸš” BREAKING : Pemain Timnas Indonesia Pratama Arhan dipastikan bakal meninggalkan Tokyo Verdy. Kepastian itu disampaikan oleh sang agen Dusan Bogdanovic. Kontrak Pratama Arhan sendiri disebut akan berakhir pada akhir tahun 2023. Arhan kabarnya akan henggang ke klub kastaâ€ https://t.co/oexkOLIQa5	['breaking', 'pemain', 'timnas', 'indonesia', 'pratama', 'arhan', 'dipastikan', 'bakal', 'meninggalkan', 'tokyo', 'verdy', 'kepastian', 'itu', 'disampaikan', 'oleh', 'sang', 'agen', 'dusan', 'bogdanovic', 'kontrak', 'pratama', 'arhan', 'sendiri', 'disebut', 'akan', 'berakhir', 'pada', 'akhir', 'tahun', 'arhan', 'kabarnya', 'akan', 'henggang', 'ke', 'klub', 'kasta']

2.3.6. Konversi ke Kalimat

Setelah proses tokenisasi, setiap token merupakan anggota tersendiri dari *list*. Untuk itu, *list* tadi perlu dikonversi ke dalam bentuk kalimat sebelum dilakukannya proses *stemming*. Berikut contoh hasil proses konversi ke kalimat pada **Tabel 8**.

Tabel 8. Konversi ke Kalimat

full_text	kalimat
ðŸš” BREAKING : Pemain Timnas Indonesia Pratama Arhan dipastikan bakal meninggalkan Tokyo Verdy. Kepastian itu disampaikan oleh sang agen Dusan Bogdanovic. Kontrak Pratama Arhan sendiri disebut akan berakhir pada akhir tahun 2023. Arhan kabarnya akan henggang ke klub kastaâ€ https://t.co/oexkOLIQa5	breaking pemain timnas indonesia pratama arhan meninggalkan tokyo verdy kepastian sang agen dusan bogdanovic kontrak pratama arhan arhan kabarnya henggang klub kasta

2.3.7. Stemming

Stemming merupakan proses untuk merubah kata berimbuhan menjadi kata dasarnya [3]. Pada proses *stemming* digunakan *library* Sastrawi yang merupakan *library* untuk NLP dalam bahasa Indonesia. Digunakan kelas *Stemmer_Factory()* untuk melakukan proses *stemming* pada data. Berikut contoh hasil *stemming* pada **Tabel 9**.

Tabel 9. Stemming

full_text	stemming
ðŸš” BREAKING : Pemain Timnas Indonesia Pratama Arhan dipastikan bakal meninggalkan Tokyo Verdy. Kepastian itu disampaikan oleh sang agen Dusan Bogdanovic. Kontrak Pratama Arhan sendiri disebut akan berakhir pada akhir tahun 2023. Arhan kabarnya akan henggang ke klub kastaâ€ https://t.co/oexkOLIQa5	breaking main timnas indonesia pratama arhan tinggal tokyo verdy pasti sang agen dus bogdanovic kontrak pratama arhan arhan kabar henggang klub kasta

2.4. Feature Extraction

Feature extraction merupakan proses pembobotan kata menjadi vektor. Pada penelitian ini, teknik *feature extraction* yang digunakan adalah TF-IDF. Tahapan yang dilakukan pada TF-IDF adalah menghitung *Document Frequency (DF)*, *Term Frequency (TF)*, dan *Inverse Document Frequency (IDF)* [3].

TF-IDF terdiri dari dua komponen yaitu *Term Frequency (TF)* dan *Inverse Document Frequency (IDF)*. TF menentukan seberapa penting sebuah kata dilihat dari seberapa sering kata itu muncul pada sebuah dokumen. Sementara pada IDF, suatu kata dianggap penting jika kata tersebut tidak terlalu sering muncul pada dokumen lain [14]. Berikut merupakan persamaan dari TF-IDF:

$$tf_{ij} = \frac{f_d(i)}{\max_{j \in d} f_d(j)} \quad (1)$$

$$idf(t, D) = \log \left(\frac{N}{df(t) + 1} \right) \quad (2)$$

2.5. Modeling

Dalam penelitian ini, peneliti menggunakan tiga variasi perbandingan data latih dan data uji. Berikut tiga perbandingan yang digunakan:

- 70 : 30
- 75 : 25
- 80 : 20

Untuk melakukan pemodelan, peneliti menggunakan bahasa pemrograman python dan *package scikit-learn*.

2.5.1. Naïve Bayes

Naïve Bayes merupakan algoritma klasifikasi yang berbasis pada teorema *Bayes* dari ilmu statistika dan probabilitas. *Naïve Bayes* dapat memprediksi probabilitas dari jumlah anggota sebuah kelas dan probabilitas dari suatu data terhadap kelas tertentu [15]. Berikut merupakan persamaan dari teorema *Bayes*.

$$1. P(A|B) = \frac{P(B|A).P(A)}{P(B)} \quad (3)$$

2.5.2. Support Vector Machine (SVM)

Tujuan utama dari algoritma *Support Vector Machine (SVM)* adalah untuk mencari *hyperplane (Decision Boundary)* dengan *margin* maksimum. *Hyperplane* merupakan pembatas atau pemisah antara dua kelas yang diklasifikasi. Banyak *hyperplane* yang dapat dibentuk untuk mengklasifikasikan kedua kelas, tetapi *hyperplane* dengan jarak maksimum akan membuat hasil klasifikasi menjadi semakin akurat. Untuk dapat memaksimalkan *margin* setiap kelas ke *hyperplane*, digunakanlah *Support Vector*. *Support Vector* merupakan *data point* yang paling dekat dengan *hyperplane*. Jika *Support Vector* berubah, maka posisi *hyperplane* juga berubah [1] [11].

2.5.3. K-Nearest Neighbor (KNN)

Algoritma *K-Nearest Neighbor (KNN)* memiliki konsep dasar bahwa data yang mirip (memiliki jarak yang dekat) cenderung memiliki kelas yang sama [16]. Adapun langkah-langkah pada algoritma *K-Nearest Neighbor* adalah [3]:

- a. Tentukan nilai k
- b. Hitung jarak data baru dengan semua data latih dengan menggunakan perhitungan *Euclidean Distance*.

$$2. \text{distance}(x, X_i) = \sqrt{\sum_{j=1}^d (x_j - X_{ij})^2} \quad (4)$$

- c. Urutkan dari jarak terdekat
- d. Periksa kelas tetangga terdekat, jika $k > 1$ maka kelasnya adalah kelas mayoritas dari para tetangga.
- e. Kelas data baru adalah kelas tetangga terdekat atau kelas mayoritas tetangga terdekat jika $k > 1$.

2.6. Evaluation

Terakhir, model yang dibuat dengan ketiga algoritma yang telah disebutkan akan dievaluasi dan hasil evaluasi akan digunakan sebagai pembanding algoritma mana yang paling efektif. *Evaluation metrics* yang digunakan pada penelitian ini adalah *accuracy*, *accuracy* merupakan *evaluation metrics* yang menentukan efektivitas model dengan menghitung berapa banyak prediksi kelas dari model yang sesuai dengan kelas sebenarnya [15]. Berikut merupakan persamaan dari akurasi:

$$\text{Akurasi} = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (5)$$

3. Hasil dan Pembahasan

Digunakan bahasa pemrograman python dalam penulisan program dari proses pengumpulan data, *labeling*, *pre-processing*, hingga *modeling*. Google Colab dimanfaatkan sebagai IDE tempat menjalankan program dengan memanfaatkan beberapa *library* dan *package built in* seperti *Pandas*, *numpy*, dan *scikit-learn*. Digunakan pula beberapa *package* dan *library* pihak ketiga seperti *Sastrawi*, *Gensim*, dll.

Setelah melakukan proses *labeling* dan *pre-processing*, dilakukan proses *modeling* terhadap data. Proses *modeling* menggunakan tiga metode yang akan dibandingkan yaitu *Naïve Bayes*, *Support Vector Machine (SVM)*, dan *K-Nearest Neighbor (KNN)*. Digunakan *scikit-learn* untuk melakukan proses *modeling* pada ketiga metode tersebut. Khusus untuk KNN, dilakukan pengujian dengan beberapa nilai k untuk mencari nilai akurasi terbesar.

3.1. Percobaan 1

Berikut adalah hasil dari percobaan pertama pada **Tabel 10**.

Tabel 10. Percobaan 1

Perbandingan data training dan data test	75:25	70:30	80:20
Akurasi <i>Naïve Bayes</i>	77,64%	77,82%	77,45%
Akurasi SVM	77,56%	77,94%	77,81%
Akurasi KNN	65,39% (k=9)	63,77% (k=9)	64,43% (k=3)

3.1.1. Hasil Perbandingan Data 75:25

Pengujian pertama dilakukan pada perbandingan data latih dan data uji sebesar 75:25, dengan perbandingan tersebut diperoleh data latih sebanyak 4091 dan data uji sebanyak 1364. Setelah melakukan proses training pada ketiga model, dihasilkan nilai akurasi *Naïve Bayes* sebesar 77,6%; akurasi *Support Vector Machine (SVM)* 77,5%; dan akurasi *K-Nearest Neighbor (KNN)* 65,3% pada k=9.

3.1.2. Hasil Perbandingan Data 70:30

Pengujian kedua dilakukan pada data dengan perbandingan data latih dan data uji 70:30, diperoleh data latih sebanyak 3818 dan data uji sebanyak 1637. Hasil nilai akurasi dari ketiga model adalah *Naïve Bayes* sebesar 77,8%; *Support Vector Machine (SVM)* sebesar 77,9%; dan *K-Nearest Neighbor (KNN)* 63,7% pada k=9.

3.1.3. Hasil Perbandingan Data 80:20

Pengujian ketiga dilakukan pada data dengan perbandingan data latih dan data uji 80:20, dari perbandingan tersebut diperoleh data latih sebanyak 4364 dan data uji sebanyak 1091. Nilai akurasi yang didapatkan oleh ketiga model adalah *Naïve Bayes* sebesar 77,4%; *Support Vector Machine (SVM)* 77,8%; dan *K-Nearest Neighbor (KNN)* sebesar 64,4% pada k=3.

3.2. Percobaan 2

Setelah selesai dengan percobaan pertama, peneliti mencoba untuk melakukan percobaan kedua. Hal ini karena setelah melakukan uji prediksi dengan kalimat baru, terdapat beberapa kata yang sebenarnya tidak penting tetapi selalu terdeteksi menjadi kelas *bullying*. Hal itu membuat peneliti memutuskan untuk melakukan *training* ulang dengan penambahan beberapa kata baru di file *stopwords*. Berikut merupakan hasil akurasi dari proses training ulang pada percobaan kedua. Berikut pada **Tabel 11** merupakan hasil dari percobaan kedua.

Tabel 11. Percobaan 2

Perbandingan data training dan data test	75:25	70:30	80:20
Akurasi <i>Naïve Bayes</i>	78,29%	77,50%	78,64%
Akurasi SVM	77,63%	76,95%	77,91%
Akurasi KNN	65,61% (k=5)	63,26% (k=7)	60,03% (k=3)

3.2.1. Hasil Perbandingan Data 75:25

Pengujian pertama pada percobaan kedua masih menggunakan perbandingan data latih dan data uji sebesar 75:25, didapatkan data latih sebanyak 4089 dan data uji sebanyak 1364. Setelah proses modeling, nilai akurasi dari ketiga metode adalah *Naïve Bayes* 78,2%; *Support Vector Machine (SVM)* 77,6%; dan *K-Nearest Neighbor (KNN)* 65,6% dengan nilai $k=5$.

3.2.2. Hasil Perbandingan Data 70:30

Pengujian kedua menggunakan perbandingan data latih dan data uji 70:30, didapatkan 3817 data latih dan 1636 data uji. Hasil nilai akurasi pada pengujian kedua ini adalah *Naïve Bayes* 77,5%; *Support Vector Machine (SVM)* 76,9%; dan *K-Nearest Neighbor (KNN)* 63,2% dengan nilai $k=7$.

3.2.3. Hasil Perbandingan Data 80:20

Pengujian ketiga menggunakan perbandingan 80:20, didapatkan data latih sebanyak 4362 dan data uji sebanyak 1091. Pada proses modeling dihasilkan nilai akurasi sebesar 78,6% untuk *Naïve Bayes*, 77,9% untuk *Support Vector Machine (SVM)*, dan 60% untuk *K-Nearest Neighbor (KNN)* dengan nilai $k=3$.

4. Kesimpulan dan Saran

Nilai akurasi yang didapat dari percobaan kedua menunjukkan bahwa dari ketiga metode yang digunakan untuk klasifikasi, *Naïve Bayes* konsisten mendapatkan nilai tertinggi di mayoritas perbandingan data latih dan data uji. Di posisi kedua terdapat *Support Vector Machine (SVM)* yang hanya memiliki selisih 1% dari nilai akurasi *Naïve Bayes*, SVM bahkan mengungguli *Naïve Bayes* di dua perbandingan pada percobaan pertama. Sementara metode *K-Nearest Neighbor (KNN)* belum mampu menandingi kedua algoritma sebelumnya, KNN hanya mampu mendapatkan akurasi dalam rentang 60-65% di ketiga perbandingan data.

Untuk lebih lengkapnya, nilai akurasi menunjukkan jika metode *Naïve Bayes* dengan perbandingan data latih dan data uji sebesar 80 : 20 mendapatkan nilai akurasi tertinggi dengan nilai 78,6%. SVM mendapatkan nilai akurasi tertinggi pada perbandingan data 80 : 20 sebesar 77,9%. Dan KNN mendapatkan nilai akurasi terbesar pada perbandingan data 75 : 25 sebesar 65,6%.

Penelitian ini masih belum memberikan hasil yang sempurna. Nilai akurasi yang didapatkan masih dapat ditingkatkan dengan melakukan proses preprocessing yang lebih baik. Terdapat peluang di masa depan untuk memperbaiki model hasil penelitian ini terutama pada tahap *pre-processing* khususnya pada bagian normalisasi. Diperlukan proses normalisasi yang jauh lebih efektif karena terdapat beberapa kata yang memiliki banyak varian bentuk.

Daftar Pustaka

- [1] R. R. Dalvi, S. Baliram Chavan, and A. Halbe, "Detecting A Twitter Cyberbullying Using Machine Learning," in *International Conference on Intelligent Computing and Control System (ICICCS 2020)*, 2020.
- [2] Z. Alsaed and D. Eleyan, "APPROACHES TO CYBERBULLYING DETECTION ON SOCIAL NETWORKS: A SURVEY," *J Theor Appl Inf Technol*, vol. 15, no. 13, 2021, [Online]. Available: www.jatit.org

-
- [3] R. M. Candra and A. N. Rozana, "Klasifikasi Komentar Bullying pada Instagram Menggunakan Metode K-Nearest Neighbor," *IT Journal Research and Development*, vol. 5, no. 1, pp. 45–52, Jul. 2020, doi: 10.25299/itjrd.2020.vol5(1).4962.
- [4] A. Akhter, K. A. Uzzal, and M. M. A. Polash, "Cyber Bullying Detection and Classification using Multinomial Naïve Bayes and Fuzzy Logic," *International Journal of Mathematical Sciences and Computing*, vol. 5, no. 4, pp. 1–12, Nov. 2019, doi: 10.5815/ijmsc.2019.04.01.
- [5] A. Desai, S. Kalaskar, O. Kumbhar, and R. Dhumal, "Cyber Bullying Detection on Social Media using Machine Learning," *ITM Web of Conferences*, vol. 40, p. 03038, 2021, doi: 10.1051/itmconf/20214003038.
- [6] K.D. Gorro, M. J. G. Sabellano, K. Gorro, C. Maderazo, and K. Capao, "Classification of Cyberbullying in Facebook Using Selenium and SVM," in *2018 3rd International Conference on Computer and Communication System (ICCCS)*, 2018, pp. 183–186.
- [7] M. Di Capua, E. Di Nardo, and A. Petrosino, "Unsupervised cyber bullying detection in social networks," in *Proceedings - International Conference on Pattern Recognition*, Institute of Electrical and Electronics Engineers Inc., Jan. 2016, pp. 432–437. doi: 10.1109/ICPR.2016.7899672.
- [8] T. de Souza Dias and S. Thapa, "Tackling Football-Related Online Hate Speech: The Role of International Human Rights Law: Part I," <https://www.ejiltalk.org/>.
- [9] A. Arimoro and A. Elgujja, "WHEN DISSENT BY FOOTBALL FANS ON SOCIAL MEDIA TURNS TO HATE: CALL FOR STRICTER MEASURES," *University of Maidguri Journal of Public Law*, vol. 6, no. 1, 2019, [Online]. Available: <https://twitter.com/RealMadrid>
- [10] D. Kilvington and J. Price, "Tackling Social Media Abuse? Critically Assessing English Football's Response to Online Racism," *Communication and Sport*, vol. 7, no. 1, pp. 64–79, Feb. 2019, doi: 10.1177/2167479517745300.
- [11] Y. Setiawan, N. Ulva Maulidevi, K. Surendro, and P. Korespondensi, "DETEKSI CYBERBULLYING DENGAN MESIN PEMBELAJARAN KLASIFIKASI (SUPERVISED LEARNING): PELUANG DAN TANTANGAN," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTik)*, vol. 9, no. 7, pp. 1577–1582, 2022, doi: 10.25126/jtiik.202296747.
- [12] M. B. Ismiati, "Deteksi Komentar Negatif di Instagram Menggunakan Algoritma Naive Bayes Classifier," in *Prosiding Seminar Nasional Sains dan Teknologi (SNST)*, 2018, pp. 243–248.
- [13] V. Nandakumar, "CYBERBULLYING REVELATION IN TWITTER DATA USING NAÏVE BAYES CLASSIFIER ALGORITHM," *International Journal of Advanced Research in Computer Science*, vol. 9, no. 1, pp. 510–513, Feb. 2018, doi: 10.26483/ijarcs.v9i1.5396.
- [14] P. Sulistiyawati *et al.*, "PREDIKSI KATA KASAR BERBAHASA INDONESIA MENGGUNAKAN MACHINE LEARNING BERBASIS MOBILE INFRASTRUCTURE," *Transmisi*, vol. 24, no. 2, pp. 55–61, May 2022, doi:
-

10.14710/transmisi.24.2.55-61.

- [15] IEEE Staff, “Presumptive Detection of Cyberbullying on Twitter through Natural Language Processing and Machine Learning in the Spanish Language,” in *IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, IEEE, 2019.
- [16] T. Srivastava, “A Complete Guide to K-Nearest Neighbors (Updated 2023),” www.analyticsvidhya.com. Accessed: Dec. 31, 2023. [Online]. Available: [https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/#:~:text=The%20K-Nearest%20Neighbor%20\(KNN,training%20dataset%20as%20a%20reference](https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/#:~:text=The%20K-Nearest%20Neighbor%20(KNN,training%20dataset%20as%20a%20reference)