JOURNAL OF SMART SYSTEM

Vol. 1 No. 1 Juli 2021

# Deteksi Cyberbullying pada Facebook Menggunakan Algoritma *K-Nearest Neighbor*
# *(Detect Cyberbullying on Facebook Using K-Nearest Neighbor Algorithm)*

Nur Fitrianingsih Hasan[1)], Vera Wati[2)]
[1)] Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Papua
[2)] Program Studi Sistem Informasi Kota Cerdas, Fakultas Teknik, Universitas Tunas Pembangunan Surakarta
E-mail: [1)]fitriahasan@umpapua.ac.id, [2)]vera.w@lecture.utp.ac.id,

**Abstrak**

Kasus penghinaan di Indonesia yang dilakukan di media sosial seperti Facebook sudah disinggung pada Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik UU ITE pasal 27 ayat 3. Maraknya penggunaan facebook yang meningkat secara cepat mengindikasi berbuat kejahatan. Jumlah like dan komentar bisa diketahui otomatis, namun seberapa besar sentimen positif dan negatif pengguna masih perlu dievaluasi. Evaluasi deteksi sentimen menjadi penting karena bertujuan untuk membantu kebijakan pihak facebook menindak lanjuti pelaku cyberbullying. Metode yang diusulkan dengan pendekatan K-NN yang menjadi bagian algoritma machine learning untuk menemukan kelas atau nilai K agar memperoleh nilai yang baik. Hasil akurasi menggunakan K-NN untuk deteksi cyberbullying pada facebook dalam mengenali sentiment positif cyberbullying perolehan akurasi tertinggi saat menggunakan 1-NN. Akurasi tertinggi pada lowercase dan uppercase influence perolehan sama 71,43%. Diikuti influence punctuation 71,40% dan akurasi pengaruh normalisasi kata dasar 70,80%. Namun waktu komputasi tercepat pengujian saat influence punctuation yaitu 0,00 pengujian lain menghasilkan 0,01.

**Kata Kunci**— Cyberbullying, Sentimen Analysis, KNN

***Abstract***

*Cases of insults in Indonesia carried out on social media such as Facebook have been mentioned in Law Number 11 of 2008 concerning Information and Electronic Transactions of the ITE Law article 27 paragraph 3. The widespread use of Facebook indicates a crime. The number of likes and comments can be known automatically, but how much positive and negative user sentiment still needs to be evaluated. Sentiment detection evaluation is important because it aims to help Facebook's policies follow up on cyberbullying perpetrators. The proposed method is the k-NN approach which is part of the Machine Learning algorithm to find the class or value of k in order to get a good score. The results of the accuracy of using k-NN for cyberbullying detection on facebook in recognizing positive sentiments of cyberbullying gain the highest accuracy when using 1-NN. The highest accuracy in the lowercase and uppercase influencce is the same as 71.43%. Followed by influence punctuation 71.40% and accuracy of base word normalization effect 70.80%. However, the fastest computational time for the test when the influence punctuation is 0.00, another test results in 0.01.*

***Keywords***— *Cyberbullying, Sentimen Analysis, KNN*

## 1. Pendahuluan

Dilansir pada databooks data kata pada tahun 2019, media sosial besutan Mark Zuckerberg menduduki peringkat tertinggi dalam 10 media sosial dengan pengguna aktif terbesar oktober 2018 [1]. Facebook memiliki pengguna teraktif dibandingkan dengan youtube, whatsapp, instagram dan media sosial lain yaitu mencapai 2,2 Miliar. Para pengguna facebook biasanya menggunakan jejaring sosial untuk berbagai keperluan, salah satunya menyampaikan ide gagasan dengan menggunakan halaman khusus pada facebook [2][3].

Keberadaan media sosial tentu saja memiliki dampak positif dan negatif, sebagai salah satu dampak negatif menggunakannya sebagai tempat *bullying* [4][5][6]. Kasus penghinaan di Indonesia yang dilakukan di media sosial seperti facebook, twitter, instagram dan aplikasi instan lain sudah disinggung pada undang-undang nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik (UU ITE). Tindakan menunjukan penghinaan terhadap orang lain tercermin pada pasal 27 ayat (3) UU ITE yang berbunyi : "Setiap orang dengan sengaja dan tanpa hak mendistribusikan dan/atau mentransmisikan dan/atau membuat dapat diaksesnya informasi elektronik dan/atau dokumen elektronik yang memiliki muatan penghinaan dan/atau pencemaran nama baik." Maraknya pengguna

facebook yang meningkat dengan pesat ini mengindikasi untuk berbuat kejahatan [7][8][9]. Hal tersebut bisa berasal dari jumlah *like* pada status atau komentar di facebook yang dapat diketahui secara otomatis, namun belum dapat mengetahui seberapa besar sentimen pengguna (pro dan kontra) dari komentar positif atau negatif [5][10][3][11]. Menurut [12] kinerja analisis sentimen dalam mengevaluasi akurasi klasifikasi tergantung pada tiga faktor yaitu fitur, jumlah fitur dan pendekatan klasifikasinya, hal tersebut untuk mengevaluasi kinerja klasifikasi algorima dan tingkat akurasinya.

Penelitian mengenai analisis sentiment dengan studi kasus pada facebook sudah dilakukan oleh [13] dengan menyelidiki surat kabar Teluk Arab memanfaatkan proses *text minning* dalam menghasilkan wilayah negara sering *posting*, masih tidak efektif karena kurangnya perluasan pengguna internet di Arab. Polaritas sentimen penggunaan facebook telah diteliti oleh [14] dengan pendekatan *Hybrid* yang digabungkan teknik berbasis leksikal menghasilkan akurasi yang cukup tinggi dan kasus yang dibahas mengenai deteksi perubahan emosional yang menerapkan metode sentbuk. Pemisahan konteks pada facebook untuk mengetahui konten bias manusia yang berdasarkan abtraksi, valensi emosional maupun ekstraksi fitur dalam mengetahui analisis sentimen juga dibahas dalam penelitian [15] menegaskan gagasan yang kuat mempengaruhi hasil metode *baseline*.

Klasifikasi kasus kejahatan juga dimanfaatkan pada media sosial lainnya. Kasus *phising* pada twitter menjadi penyebaran yang sulit terdeteksi dibanding *email*, namun peringatan keamanan memanfaatkan dengan teknik *Random Forest* (RF) dalam penelitian [16] menghasilkan peringatan yang efektif. Deteksi *cyberbullying* pada klasfifikasi gambar dan teks dan pendeteksian pada komentar postingan instagram menemukan bahwa teks menjadi prediktor kuat di masa depan pada gambar tertentu [17]. Pada penelitian [18] pada *social networking sites* (SNS) mengklasifikasi bullying seperti *flaming*, pelecehan, rasisme dan terorisme menggunakan algoritma genetika dan *fuzzy* pada *dataset Myspace* dan Formingspring.me. Pemanfaatan metode *Support Vector Machine* (SVM*)* dan *Hybrid Partical Swarm Optimization* (PSO) dibahas oleh [19] pada twitter untuk beropini klasifikasi positif atau negatif dalam meninjau film. Klasifikasi sentimen resepsi konsumen terhadap jaringan transportasi Uber yang dikumpulkan pada *tweets* juga dibahas dalam penelitian [20] dengan kelompok semantik yang koheren.

Selain itu pemanfaatan *k-Neareast Neighbor* maupun yang dikombinasikan dengan algoritma lain dimanfaatkan dalam penelitian, seperti halnya oleh [21] menggunakan kombinasi KNN dan *Naive Bayes* memanfaatkan fitur-fitur mencari emosi, *smiley* yang diekstrak dari sentimen tweeter. Penggunaan KNN seperti penelitian oleh [22] juga cukup baik dalam menganalisa komentar facebook untuk klasifikasi sentimen sikap seseorang pada ekspedisi barang. Algoritma KNN direkomendasikan dalam penyelesaian layak dan tidaknya dalam penerimaan bantuan dana desa yang dilakukan oleh peneliti [23] pengujian dilakukan dengan *rapidminer* dengan berbagai nilai k. Data dalam obrolan dalam *game* menjadi studi kasus yang diteliti oleh [24] karena data tersaji secara *realtime* dengan memanfaatkan fitur *query database* SQL yang dibandingkan dengan klasifikasi AI untuk pendeteksian rasis para pemain.

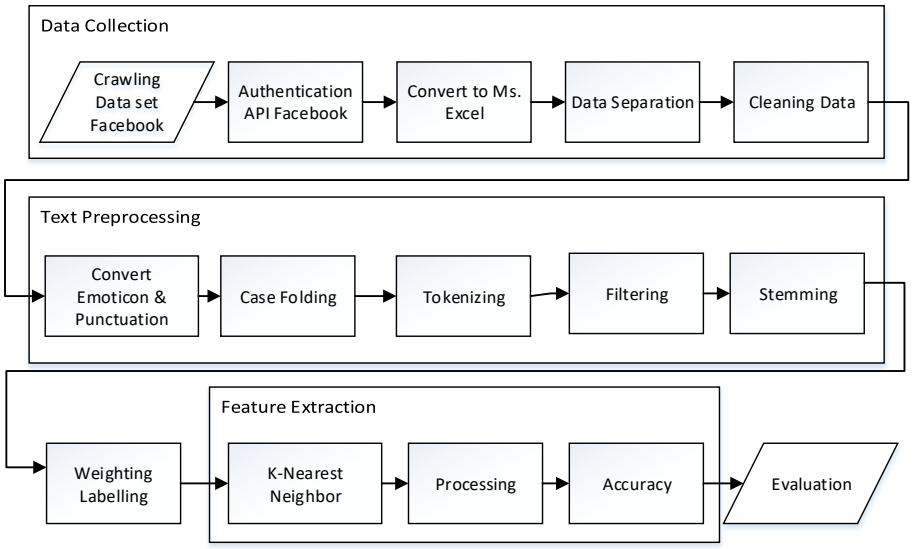
Perbandingan metode antara *k-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM) dan *Random Forest* (FR) dikenal menghasilkan akurasi tinggi, hal ini dibuktikan pada penelitian [25] pada kasus klasifikasi lahan dengan hasil *overall accuracy* (OA) tinggi dengan sampel cukup besar. Pembelajaran mesin untuk meningkatkan akurasi dengan menerapkan metode *chi-square* dan memperhitungkan fitur kata dan emotikon menghasilkan akurasi klasifikasi, akurasi menggunakan beberapa metode dan salah satunya penerapan dengan KNN pada *dataset* instagram dan twitter bahasa Turki [26]. Algoritma KNN untuk analisis sentimen data *microblog* dengan melibatkan klasifikasi KNN dan Neighbor-Weighted K-Nearest Neighbor (NW-KNN) menghasilkan akurasi yang berbeda anatara keduanya, hal ini berdasarkan pengubahan nilai k [27].

Pada penelitian terdahulu membuktikan jika media sosial bisa memiliki peluang untuk melakukan kejahatan, dengan memanfaatkan algoritma dari KNN dengan studi kasus komentar Bahasa Indonesia pada facebook diharapkan akan dapat diketahui tingkat akurasi deteksi *cyberbullying* dengan tahapan *text processing* dan penggunaan algoritma KNN.

Pada penelitian yang dilakukan, dibutuhkan klasifikasi sentimen untuk dapat mengevaluasi komentar yang mengandung *cyberbullying* bertujuan untuk membantu kebijakan pihak media sosial facebook untuk mengambil keputusan kepada penggunanya. Deteksi *cyberbulying* diberlakukan pada media sosial facebook dengan menggunakan teks berbahasa Indonesia, kemudian menerapkan metode *machine learning* KNN untuk mengetahui hasil akurasi. Pengujian akurasi akan dilakukan dengan hasil proses hubungan antara pengaruh kata yang mengandung huruf besar, huruf kecil, tanda baca dan normalisasi kata dasar pada proses pengujian sehingga didapatkan hasil yang memiliki akurasi dan komputasi optimal. Analisis sentimen dibagi dalam 2 deteksi yaitu sentimen positif mengandung kalimat *bullying* dan sentiment negatif tidak mengandung kalimat *bullying*.

## 2. Metode Penelitian

Penyelesaian penelitian digunakan beberapa tahapan untuk mempermudah pembuatan sistem dan pengujian akurasinya, seperti tampak pada Gambar 1.

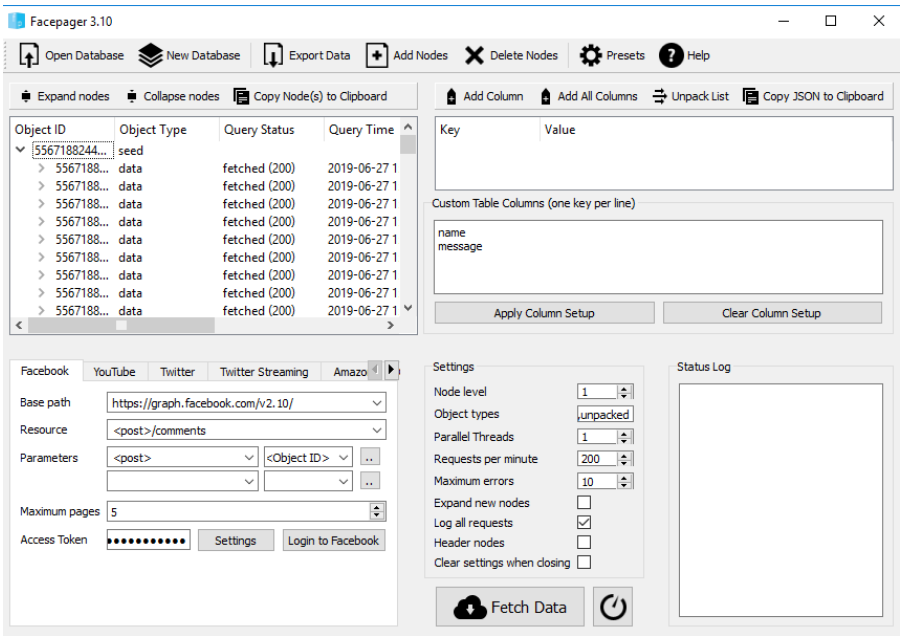



**Gambar 1.** Metode Penelitian Analisis Sentimen

### a. Data Collection

1) Crawling Dataset Facebook

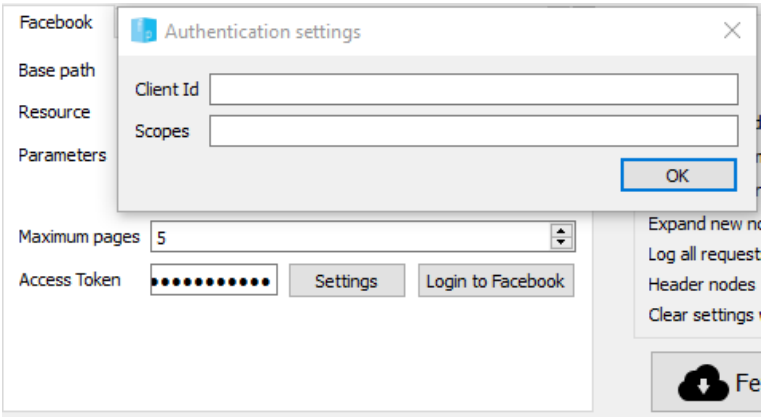
Pada tahap ini cara menghasilkan data yang dapat disimpan sebagai dataset awal yang dibutuhkan pada laman facebook [13]. Kegiatan ini mesin pencarian yang mampu mengurai dan mengekstrak data hingga tersimpan dan menghasilkan data dalam jumlah yang besar yang dibutuhkan dengan pengambilan kinerja yang tinggi [13][14] [15][16]. Proses ini mengekstrak akun *fanpage* dari profil selebriti Indonesia dengan nama akun "Deddy Corbuzier". Laman facebook dipilih peneliti karena banyaknya kunjungan dan menampung banyak komentar dari pengguna facebook lainnya. *Crawling* menggunakan Facepager sebuah platform API *open source* yang dapat mengumpulkan data secara cepat dan relevan [17][18]. Kegiatan tersebut terdapat pada Gambar 2.




**Gambar 2.** Crawling Data dari Halaman Facebook

2) Authentication API Facebook

Setelah didapatkan data mentah pada proses *crawling*, kemudian autentikasi API dari url *fanpage* profil yang diambil dari ID *user* akun, proses akses token bisa dilihat pada Gambar 3.




**Gambar 3.** Authentication API Facebook

3) Convert to file.csv

Jika data sudah berhasil didapatkan maka akan tersimpan pada *database*, kemudian dilakukan *export* dalam format .csv. Hal ini untuk mempermudah proses pengolahan data yaitu dilakukan pembersihan data yang tidak perlu dalam penelitian.

4) Data Separation

Pada tahap ini adalah pengolahan data yaitu mengelola dengan memilih kolom informasi yang penting. Selain itu membatasi jumlah dataset menjadi 3000 unggahan komentar, dimana awalnya mendapatkan data mentah 9095 postingan.

5) Data Cleaning

Setelah data siap, maka dilakukan proses *data cleaning* yaitu teknik untuk penyusutan dan menyimpulkan nilai data yang akan diolah dalam *text processing* [19][20][21]. Metode dalam pembersihan data ini dengan mempertimbangkan variabel berupa fitur komentar. Tampilan hasil *cleaning* terdapat pada Gambar 4.

| Komen |
|---|
| Always good day with underarmourid being their Brand Ambassador for two years now Sampe toko2 OL yg jual baju palsu nya juga pake foto gue but its okay Kalau memang jadi |
| Ada yg dari jauh ngambil selfie Ga dpt Gue samperin tiba tiba KAGET this what I love in my show Doing Fun shit on the middle of it officialhitamputihtrans7 |
| Back to work out Just done taking pictures for underarmourid corbuziergymosbond dan gw keram bukan krn latian tapi krn hrs diam Dia blg tahan mas Tahan Tahan Gue Keram |
| Saying thank you to all of you brothers and sisters For support and all blessing And also to this guy here Irjen Pol Drs H Ahmad Dofiri MSi Who willing to spent time to witness m |
| gusmiftah ga kuat Ini harus di posting HAHAHAHHAAHHAAAHAA Artis Korea Pak Pak Panjenengan KAREP mandi susu Pak |
| Baru di ikirn foto ini Kok saya mirip Bodyguard beliau jadinya khmarufamin sekali lagi terimakasih untuk sholat pertama nya makan malam nya bincang2 nya Just info Beliau san |
| khmarufamin Masha Allah nikmat nya hanya dari sebuah kesederhanaan |
| Alhamdulillah Kyai Said Aqil Siradj |
| Alhamdulillah |
| Ada yg tau nih anak siapa Damn his good |
| noagusnoparty netizen prank |
| agunghercules88 LOE HARUS SEMBUH BRO YANG DOAIN BANYAK GUE GA MAU TAU POKOKNYA LOE HARUS BISA BANGUN LAGI DR SANA ADU PANCO LAGI SAMA GUE I BELIEVE U |
| Doing 120 KG incline while my Mom Walking around on my back Supeeb corbuziergymosbond chestworkout gymmotivation |
| Cant stop laughing at THIS SHIT HERE HAHAHAHA And no this isnt bullying nor hurting anyone at all We used head protection for ricocepero but still this is hilarious martialart ka |
| Moment that I always love being part of this amazing show in more than 8 years Happiness is when what you think what you say and what you do are in harmony Do u? Love u a |
| PARTNER TUH GINI vagustiant |
| Ngobrol sama Menhub budikaryas bahas apa? Gimana cara gue Mundurin jam tangan beliau tanpa di sentuh Its Magic Ga kerasa lama ngobrol nya Krn jam nya mundur trus |
| EFEK LIBUR KELAMAAN SOK BUATIN CAPTION LAGI SONO |
| Tolong Di kasih CAPTION Yg kreatif |
| Bila Idul Fitri adalah lentera izinkanlah aku untuk membuka tabirnya dengan maaf agar cahayanya dapat menembus jiwa yang fitrah dari tiap khilaf Selamat Hari Raya Idul Fitri 14 |
| On YouTube now Love doing it Now we share it |
| Happy Birthday azkacorbuzier One thing I know is everyone love u Tau dia minta apa? Tiup lilin Doang Not perfect but amazing |
| Malam ini azkacorbuzier ultah iam 12 malam I love u son Be a great kid Sorry if Im not perfect as a Father but Im doing my best Now LETS FIGHT ME IN BOXING |

**Gambar 4.** Hasil dari Cleaning Data

**b. Text Preprocessing**

1) Convert Emoticon dan Punctuation

*Preprocessing* diawali dengan *convert* emotikon dan tanda baca yang memiliki makna. Misal pengubahan yang terjadi emoticon *"smiley"* yang diganti dengan kalimat bahagia dan tanda baca "&" disetarakan dengan huruf "dan" serta lainnya.

2) Case Folding
Pada tahap ini dilakukan pengubahan teks agar konsisten dengan konversi huruf dalam bentuk standar [22][23]. Pengubahan ini dilakukan agar *user* dapat menerima informasi dengan benar yaitu mengubah huruf kecil dan hanya menampung karakter abjad "a" sampai "z" dalam penyusunan katanya, karakter selain itu dihilangkan. Misal ketidak konsistenan penulisan "InfORmatioN", "infrmatin" maka tetap memilliki makna dan hasil yang sama yaitu "information".

3) Tokenizing
*Tokenizing* melakukan pemecahan dan pemotongan kata pada teks yang terdapat pada string input, pemisah ini menggunakan tanda baca yang berperan sebagai pemisah [23][24]. Ilustrasi pada Gambar 5.

| Comment | Tokenize |
|---|---|
| mamah papahku | mamah papahku |
| Bingung, suka pusing cewek kok berjakun. | bingung suka pusing cewek kok berjakun |
| Semoga, tersadar ya. | semoga tersadar ya |
| Pasangan serasi. Dari pada menyukaimu | pasangan serasi dari pada menyukaimu |

**Gambar 5.** Tokenizing Process

4) Filtering
Proses ini menekankan pada membuang *stoplist* yaitu kata penghubung dalam bahasa Indeonesia seperti "di, dari, yang" dan lainnya yang tidak deskriptif dan menyimpan wordlist atau kata yang penting [25]. Ilustrasi pada Gambar 6.

| Tokenize | Filtering |
|---|---|
| mamah papahku | mamah papah |
| bingung suka pusing cewek kok berjakun | bingung suka cewek berjakun |
| semoga tersadar ya | semoga tersadar |
| pasangan serasi dari pada menyukaimu | pasangan serasi menyukaimu |

**Gambar 6.** Filtering Process

5) Stemming
Teknik ini diperlukan agar teks dapat dikenali, maka dilakukan pengelompokan membuang imbuhan pada kata hingga didapatkan kata dasar [26]. Ilustrasi pada Gambar 7.

| Filtering | Stemming |
|---|---|
| mamah papah | mamah papah |
| bingung suka cewek berjakun | bingung suka cewek jakun |
| semoga tersadar | semoga sadar |
| pasangan serasi menyukaimu | pasangan serasi suka |

**Gambar 7**. Stemming Process

6) Weighting Labelling
Jika file sudah dilakukan tahapan *text processing*, maka selanjutnya dilakukan tahapan *labelling* yaitu dengan memberikan pembobotan berupa klasifikasi sentimen label positif mengandung kalimat *bullying* dan negatif atau tidak mengandung kalimat *bullying*. Klasifikasi sentimen pada peneltian ini menggunakan *dataset* unggahan maupun komentar sebanyak 3000 data.

**c. Featured Extraction**
1) K-Nearest Neighbor Algorithm (Algoritma KNN)
K-Nearest Neighbor (KNN) merupakan algoritma *machine learning* pengklasifikasian untuk menemukan kelas atau nilai k yang optimal memperoleh hasil yang lebih baik [27]. Pengujian algoritma menggunakan aplikasi WEKA. Perhitungan KNN memperhatikan antara titik yang dikenal dan titik perkiran data k terdekat, dalam memilih $(y_1, y_2, \dots \dots . y_k)$ titik $y_1$ titik paling dekat dengan titik perkiraan, $y_2$ titik terdekat ke 2 dan dilanjutkan sampai ditemukan $y_k$ optimal, rumus persamaan sebagai berikut [28]

$$S(i) = \frac{1}{k} x \sum_{j=1}^{k} (s_{yj}) \quad (1)$$

Dimana;
$S(i)$ = mewakili nilai perkiraan
$k$ = kelas titik perkiraan
$j$ = perkiraan nilai rata-rata ($j = 1,2, \ldots . k$
$s_{yj}$ = nilai rata-rata yang mewakili titik data terdekat $ke -_j$

2) Processing
Proses untuk mengevaluasi kinerja dari pemodelan sistem ini, pengolahan ini dilakukan dengan *cross-validation* (CV) yaitu suatu metode statistik dari dua subset kemudian data divalidasi sehingga hasil proses pembelajaran tersebut dapat dievaluasi [11][29]. *Prosessing* terbagi menjadi 2 yaitu data *training* dan data *testing*, lihat Tabel 1.

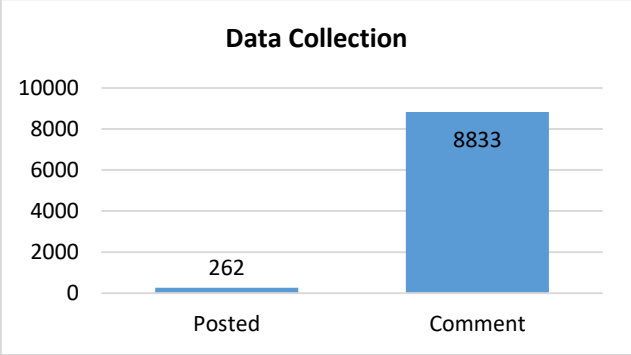
3) Accuracy
Akurasi menjadi ukuran metode dan kinerja dari sistem, yaitu menjadi pengujian untuk mengetahui hasil proses tingkat kedekatan antara nilai aktual dengan nilai prediksinya [30]. Uji akurasi dilakukan untuk mencari akurasi dengan komputasi yang tertinggi dan digunakan untuk 10-fold CV.

4) Evaluations
Tahap akhir adalah evaluasi dengan dilakukannya 4 skenario pengujian yaitu sebelum diberlakukan *text processing* maka dilakukan perbandingan dengan pengaruh huruf besar dan huruf kecil dalam akurasi, tingkat pengaruh tanda baca pada teks dan pengaruh normalisasi terhadap klasifikasi deteksi *cyberbullying*.

## 3. Hasil dan Pembahasan

a. Hasil Data Collection

Proses *data collection* untuk mengambil data postingan dan komentar pada *fanpage* dari profil selebriti Indonesia dengan nama akun “Deddy Corbuzier” selama 5 bulan terkahir mendapatkan sebanyak 9095 data.

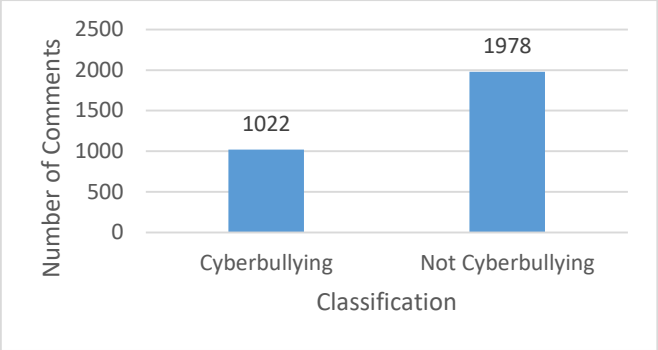



**Gambar 8**. Data Collection Posting dan Komentar

Pada Gambar 8 hasil dari pengambilan data yaitu sebanyak total 9095 data collection terdiri dari 262 data postingan dan 8.833 data komentar. *Data collection* disimpan dalam bentuk file .csv yang selanjutnya akan diolah pada tahap *preprocessing* data dan menghasilkan klasfikasi terhadapat komentar mupun postingan.

b. Hasil Klasifikasi

Setelah dilakukan tahap *preprocessing* pada tahapan yang telah dijlaskan pada metode penelitian yaitu dengan menghilangkan data kosong, teks berbahasa inggris, dll kemudian data set yang diammbil menjadi 3000 data yang kemudian dilakukan klasifikasi terhadap 3000 data komentar dan postingan ini. Pada Gambar 9 hasil klasifikasi sentimen pada dataset dari 3000 data tersebut didapatkan 1022 komentar dengan sentimen positif *cyberbullying* dan 1978 sentimen negatif *cyberbullying*.




**Gambar 9**. Klasifikasi Komentar

c. Hasil Evaluasi

Setelah mendapatkan dataset dan hasil klasifikasi maka selanjutnya dilakukan pengujian dengan menggunakan metode algoritma KNN. Pada proses *stemming* adalah normalisasi komentar sehingga mengurangi jumlah variasi kata yang memiliki makna yang sama, dengan dilakukannya stemming diharapkan dapat meningkatkan akurasi dan mempermudah untuk melakukan analisis. Evaluasi dilakukan dengan pengaruh huruf besar dan kecil, yaitu pemrosesan yang dilakukan dengan menguji akurasi tingkat pengaruh dalam menyertakan kualifikasi tersebut. Selain itu, pengaruh tanda baca juga dlakukan evaluasi serta setelah dilakukan normalisasi kata dasar dilakukan pengujian ulang. Tentu saja hal tersebut dilakukan untuk memperoleh akurasi yang tertinggi.

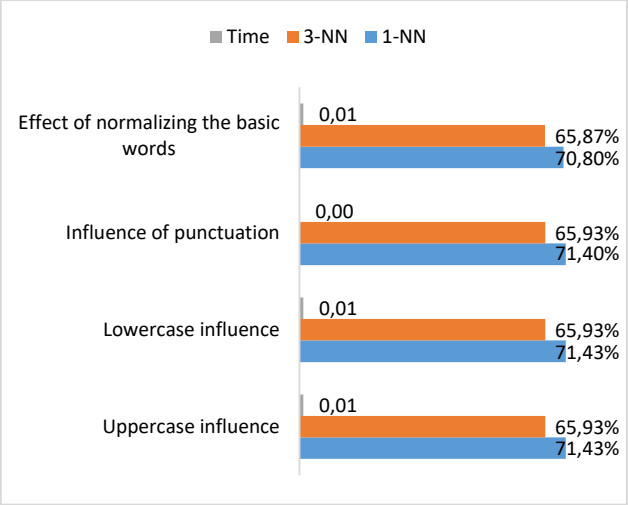
Selanjutnya setelah proses *stemming* dilakukan pengujian data dilakukan dengan metode K-Fold *Cross-validation*, pada penelitian ini nilai k menggunakan nilai 10. Tahap pengujian ini data *training* dan data *testing* di bagi seperti pada tabel 1.

**Tabel 1.** Data 10 Cross Validation

| Fold | Data Training | Data Testing |
|---|---|---|
| 1 | 301-3000 | 1-300 |
| 2 | 1-300, 600-3000 | 301-600 |
| 3 | 1-600, 900-3000 | 601-900 |
| 4 | 1-900, 1200-3000 | 901-1200 |
| 5 | 1-1200, 1500-3000 | 1201-1500 |
| 6 | 1-1500, 1800-3000 | 1501-1800 |
| 7 | 1-1800, 2000-3000 | 1801-2000 |
| 8 | 1-2100, 2400-3000 | 2101-2400 |
| 9 | 1-2400, 2700-3000 | 2401-2700 |
| 10 | 1-2700 | 2701-3000 |

Pada pengujian dengan 10-CV dimana data training sebelumnya yang telah dilatih (*training*) kemudian menjadi *data testing* untuk data selanjutnya, begitu seterusnya hingga data terakhir. Setelah dihasilkan *data testing* maka dilakukan pengujian akurasi, hasil terdapat pada Gambar 10.

Pengujian dengan normalisasi kata dasar menghasilkan akurasi tertinggi 70,80% saat menggunakan nilai 1-NN dibanding 3-NN. Sama halnya dengan pengujian yang lain, lebih tinggi peroleh ketika menggunakan 3-NN. Pengaruh *influence punctuation* mendapat akurasi 71,40% dan testing penggunaan huruf besar dan huruf kecil menghasilkan nilai akurasi yang sama yaitu 71,43%. Lama komputasi pengujian dengan 4 kualifikasi tersebut rata-rata adalah 0,01 *persecond*, kecuali pengaruh *influence* puntuation memiliki waktu komputasi paling rendah yaitu 0,00




**Gambar 10.** Hasil Akurasi

## 4. Kesimpulan

Pada penelitian yang telah dilakukan, dapat disimpulkan bahwa penggunaan algoritma KNN dalam pendeteksi *cyberbullying* di facebook menghasilkan tingkat akurasi tertinggi untuk mendeteksi sentimen positif mengandung *cyberbulling* yaitu saat menggunakan 1-NN dengan hasil akurasi pengaruh normalisasi kata dasar 70,80%, *influence punctuation* 71,40% dan perolehan akurasi paling tertinggi saat *lowercase* dan *uppercase influencce* yaitu memperoleh hasil yang sama 71,43%. Jika menggunakan 3-NN hasil akurasi tertinggi bernilai 65,93% pada tiga kualifikasi yaitu *influence of punctuation, lowercase* dan *uppercase influence* kemudian diikuti pengaruh normalisasi kata dasar perolehan 65,87%. Namun waktu komputasi tercepat pengujian saat *influence punctuation* yaitu 0,00 pengujian lain menghasilkan 0,01.

Referensi

[1] "Data Kata." .

[2] R. C. Antonius and Y. Lukito, "Klasifikasi Sentimen Komentar Politik dari Facebook Page Menggunakan Naive Bayes," *JUISI*, vol. 02, no. 02, pp. 26–34, 2016.

[3] W. Kaur, V. Balakrishnan, O. Rana, and A. Sinniah, "Liking, sharing, commenting and reacting on Facebook: User behaviors' impact on sentiment intensity," *Telemat. Informatics*, vol. 39, pp. 25–36, 2019, doi: 10.1016/j.tele.2018.12.005.

[4] R. Damayanti, "Penggunaan Bahasa Alay pada Bullying Anak di Media Sosial," *J. Autentik*, vol. Vol. 1, no. No. 2, pp. 1–11, 2017.

[5] M. N. Yusoff, A. Dehghantanha, and R. Mahmod, "Forensic Investigation of Social Media and Instant Messaging Services in Firefox OS: Facebook, Twitter, Google+, Telegram, OpenWapp, and Line as Case Studies," *Contemp. Digit. Forensic Investig. Cloud Mob. Appl.*, pp. 41–62, Jan. 2017, doi: 10.1016/B978-0-12-805303-4.00004-6.

[6] F. Del Vigna, A. Cimino, F. Dell'Orletta, M. Petrocchi, and M. Tesconi, "Hate Me, Hate Me Not: Hate Speech Detection on Facebook," *CEUR Workshop Proc.*, vol. 1816, pp. 86–95, 2017.

[7] F. Poecze, C. Ebster, and C. Strauss, "Social media metrics and sentiment analysis to evaluate the effectiveness of social media posts," *Procedia Comput. Sci.*, vol. 130, pp. 660–666, Jan. 2018, doi: 10.1016/J.PROCS.2018.04.117.

[8] S. T. Aroyehun and A. Gelbukh, "Aggression Detection in Social Media: Using Deep Neural Networks,

Data Augmentation, and Pseudo Labeling," *Proc. First Work. Trolling, Aggress. Cyberbullying*, pp. 90–97, 2018.

[9] F. N. Dezfouli, A. Dehghantanha, B. Eterovic-Soric, and K.-K. R. Choo, "Investigating Social Networking applications on smartphones detecting Facebook, Twitter, LinkedIn and Google+ artefacts on Android and iOS platforms," *Aust. J. Forensic Sci.*, vol. 48, no. 4, pp. 469–488, 2016, doi: 10.1080/00450618.2015.1066854.

[10] L. Herlina, "Disintegrasi Sosial dalam Konten Media Sosial Facebook," *J. Pembang. Sos.*, vol. 1, no. No. 2, pp. 232–258, 2018.

[11] M. Meire, M. Ballings, and D. Van den Poel, "The Added Value of Auxiliary Data in Sentiment Analysis of Facebook Posts," *Decis. Support Syst.*, vol. 89, pp. 98–112, Sep. 2016, doi: 10.1016/J.DSS.2016.06.013.

[12] A. Al-saffar, S. Awang, H. Tao, N. Omar, W. Al-saiagh, and M. Al-bared, "Malay Sentiment Analysis Based on Combined Classification Approaches and Senti-Lexicon Algorithm," *PLoS One*, pp. 1–18, 2018, doi: 10.13140/RG.2.2.33420.72320.

[13] S. Khalil and M. Fakir, "RCrawler: An R package for parallel web crawling and scraping," *SoftwareX*, vol. 6. pp. 98–106, 2017, doi: 10.1016/j.softx.2017.04.004.

[14] S. A. Catanese, P. De Meo, E. Ferrara, G. Fiumara, and A. Provetti, "Crawling Facebook for Social Network Analysis Purposes," in *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*, 2011, pp. 52:1--52:8, doi: 10.1145/1988688.1988749.

[15] M. Kumar, A. Bindal, R. Gautam, and R. Bhatia, "Keyword query based focused Web crawler," 2018, doi: 10.1016/j.procs.2017.12.075.

[16] K. H. Kim *et al.*, "A text-based data mining and toxicity prediction modeling system for a clinical decision support in radiation oncology: A preliminary study," *J. Korean Phys. Soc.*, vol. 71, no. 4, pp. 231–237, 2017, doi: 10.3938/jkps.71.231.

[17] A. Vierecke, "Using social media data for science: Till Keyling on 'Facepager,'" *alumniportal-deutschland*, 2014. .

[18] T. Keyling and J. Jünger, "Facepager. An application for generic data retrieval through APIs," *Source code and releases available at https://github. com/strohne/Facepager/(last accessed 4 May 2017)*, 2017. .

[19] J. Rammelaere and F. Geerts, "Cleaning data with forbidden itemsets," *IEEE Trans. Knowl. Data Eng.*, 2019.

[20] C. Mayfield, J. Neville, and S. Prabhakar, "ERACER: A Database Approach for Statistical Inference and Data Cleaning," in *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, 2010, pp. 75–86, doi: 10.1145/1807167.1807178.

[21] P. Li, X. Rao, J. Blase, Y. Zhang, X. Chu, and C. Zhang, "CleanML: A Benchmark for Joint Data Cleaning and Machine Learning [Experiments and Analysis]," 2019.

[22] Y.-H. Tseng, C.-J. Lin, and Y.-I. Lin, "Text mining techniques for patent analysis," *Inf. Process. Manag.*, vol. 43, no. 5, pp. 1216–1247, Sep. 2007, doi: 10.1016/J.IPM.2006.11.011.

[23] S. Vijayarani, M. J. Ilamathi, and M. Nithya, "Preprocessing techniques for text mining-an overview," *Int. J. Comput. Sci. Commun. Networks*, vol. 5, no. 1, pp. 7–16, 2015.

[24] R. Duwairi and M. El-Orfali, "A study of the effects of preprocessing strategies on sentiment analysis for Arabic text," *J. Inf. Sci.*, vol. 40, no. 4, pp. 501–513, Aug. 2014, doi: 10.1177/0165551514534143.

[25] "Indonesian Wordlist," 2019. .

[26] A. Yulio, “Steeming Bahasa Indonesia dengan Python Sastrawi,” 2017. .

[27] R. Agrawal, “Integrated Parallel K-Nearest Neighbor Algorithm,” 2019, pp. 479–486.

[28] G.-F. Fan, Y.-H. Guo, J.-M. Zheng, and W.-C. Hong, “Application of the Weighted K-Nearest Neighbor Algorithm for Short-Term Load Forecasting,” *Energies*, vol. 12, no. 5, 2019, doi: 10.3390/en12050916.

[29] R. L. Hasanah, M. Hasan, W. E. Pangesti, W. Gata, and F. F. Wati, “Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode Knn ( K-Nearest Neighbor ),” *J. TECHNO Nusa Mandiri*, vol. 16, no. 1, pp. 1–6, 2019.

[30] T. J. Lee, J. Gottschlich, N. Tatbul, E. Metcalf, and S. Zdonik, “Precision and recall for range-based anomaly detection,” *arXiv Prepr. arXiv1801.03175*, 2018.