Blend Sains
Jurnal Teknik
https://jurnal.ilmubersama.com/index.php/blendsains

Artikel Penelitian (Teknik Informatika)

Klasifikasi Akun *Buzzer* Menggunakan Algoritma *K-Nearest Neighbor* pada Tagar #STYTanpaDiasporaNol di Media Sosial X

Afiq Alghazali Lubis [1], Said Iskandar Al Idrus [1], Zulfahmi Indra [1], Kana Saputra S [1], Chairunisah [2]*

[1] *Fakultas Matematika dan Ilmu Pengetahuan Alam, Program Studi Ilmu Komputer, Universitas Negeri Medan, Medan, Indonesia*
[2] *Fakultas Matematika dan Ilmu Pengetahuan Alam, Program Studi Matematika, Universitas Negeri Medan, Medan, Indonesia*

INFORMASI ARTIKEL

Diterima Redaksi: 16 Juli 2025
Revisi Akhir: 16 Oktober 2025
Diterbitkan *Online*: 19 Oktober 2025

KATA KUNCI

Buzzer
Klasifikasi
Sosial Media

KORESPONDENSI (*)

Phone: +62 822-7346-3107
E-mail: afiqalghazali.edu@gmail.com

A B S T R A K

Peningkatan pengguna media sosial X pada tahun 2024 sebesar 639 ribu mengakibatkan penyebaran informasi yang sangat masif, menjadikan *buzzer* berperan dalam mengarahkan opini publik dan memicu konflik sosial, seperti yang terlihat pada tren #STYTanpaDiasporaNol usai gugurnya tim nasional Indonesia di ASEAN *Championship* 2024. Penelitian ini bertujuan untuk membangun model *machine learning* dalam klasifikasi akun *buzzer* menggunakan algoritma *K-Nearest Neighbor* (KNN). Data yang akan digunakan dalam penelitian ini berasal dari kumpulan *tweet* dari sosial media X dalam tagar #STYTanpaDiasporaNol. Penelitian ini memiliki prosedur penelitian, di antaranya pengumpulan data, pra-pemrosesan data (*cleaning*, *labelling*, *feature engineering* dan *standardization*), *splitting* data, pemrosesan data, dan evaluasi model. Hasil penelitian ini mendapatkan model dengan akurasi terbaik yaitu varian model perbandingan *split* data 80:20 dan K = 5 dengan nilai akurasi sebesar 89% serta nilai *precision* dan *recall* sebesar 89% lalu nilai *F1-score* sebesar 88%. Model sangat baik dalam memprediksi kelas mayoritas namun kesulitan dalam memprediksi kelas minoritas. Kemudian dilakukan eksperimen *resampling* data dengan tujuan membuat keseimbangan data. Hasil didapatkan bahwa varian pada *split* data 70:30 dengan K = 9 diperoleh akurasi sebesar 91% dengan *precision*, *recall* dan *accuracy* juga sebesar 91%. Model eksperimen ini cukup baik mendeteksi kelas mayoritas maupun kelas minoritas.

PENDAHULUAN

Perkembangan media sosial saat ini sudah memudahkan pengguna internet untuk berkomunikasi secara langsung dan penyebaran informasi yang cukup masif. Merujuk data yang dikutip dari We Are Social, menunjukkan bahwa sebanyak 139 juta pengguna media sosial di Indonesia per Januari 2024 dan apabila dibandingkan dengan total populasi di Indonesia berada di angka 49,9% [1]. Salah satu media sosial yang cukup digemari, yaitu X di Indonesia pada awal tahun 2024 mengalami peningkatan sebesar 693 ribu pengguna dibandingkan dengan tahun 2023 [1]. Dengan algoritma yang dimiliki oleh X, dapat merangkum hal yang lagi ramai diperbincangkan dalam fitur *trending topic*. Hal yang sedang *trending* umumnya diikuti oleh tagar atau *hashtag* yang berisi topik pembahasan.

Cepatnya penyebaran informasi ini dimanfaatkan oleh *buzzer*, yang awalnya merupakan sebuah teknik *marketing*, kini menjadi alat propaganda publik [2]. *Buzzer* dapat bekerja secara personal maupun kolektif dalam mengangkat atau mendengungkan (*buzzing*) suatu isu agar dibicarakan oleh publik dalam dunia maya [3]. Keberadaan *buzzer* di antara masyarakat berisiko menjadi penyebab perpecahan dan konflik sosial di negara Indonesia. Salah satu konflik sosial yang terjadi disebabkan oleh *buzzer* yaitu pada Pemilihan Presiden tahun 2019 ditandai dengan adanya perang komentar kebencian serta munculnya istilah “Cebong” dan “Kampret” untuk masing-masing kubu [4].

https://doi.org/10.56211/blendsains.v4i2.1093

Adapun konflik sosial lainnya yang terdapat campur tangan *buzzer* di luar dari kon*test*asi politik adalah munculnya tagar #STYTanpaDiasporaNol pada hari Minggu, tanggal 22 Desember 2024, tepat sehari setelah kekalahan tim nasional Indonesia melawan tim nasional Filipina dengan skor 0-1 yang sekaligus menandakan gugurnya tim nasional Indonesia dari ajang ASEAN *Championship* 2024 [5]. Hal ini dibuktikan oleh sebuah tangkapan layar percakapan tentang penawaran untuk menjadi *buzzer* serta instruksi dalam melancarkan pesan provokatif [6].

Munculnya tagar ini sebagai *trending topic* memicu perdebatan dan perpecahan dalam pendukung Timnas yang ditandai dengan munculnya tagar dari kedua belah pihak yaitu #STYStay dan #STYOut [7]. Masa depan karier Shin Tae-Yong sebagai pelatih Timnas Indonesia ditentukan pada konferensi pers yang dilakukan oleh PSSI pada tanggal 6 Januari 2025 yang menyatakan pemutusan kontrak sebagai pelatih Timnas Indonesia. Keputusan ini menimbulkan rasa kekecewaan pendukung Timnas Indonesia atas langkah yang dilakukan PSSI, dikarenakan peningkatan capaian dan prestasi yang didapatkan oleh Timnas Indonesia, terutama dalam 15 tahun terakhir [8]. Penekanan transparansi dalam keputusan PSSI ini dilakukan oleh beberapa pihak, salah satunya Komisi X DPR RI yang menanyakan hasil evaluasi kinerja secara objektif serta mempertimbangkan aspirasi publik [9]. Efek yang diciptakan oleh propaganda *buzzer* pada tagar #STYTanpaDiasporaNol menciptakan polarisasi publik serta keraguan transparansi dalam lembaga PSSI oleh masyarakat.

Berdasarkan dari dampak negatif yang disebabkan oleh akun *buzzer*, proses pengklasifikasian akun *buzzer* dibutuhkan untuk membedakan akun yang terindikasi sebagai media provokator yang membawa informasi tanpa validasi dengan akun yang hanya menyampaikan opini pribadi saja. Sebelumnya, penelitian dalam klasifikasi akun *buzzer* telah dilakukan dengan metode yang berbeda. Algoritma *Decision Tree* C4.5 digunakan dalam menyelesaikan permasalahan dengan hasil rata-rata akurasi yang diperolah adalah 88,5% [10]. Klasifikasi akun *buzzer* juga telah dilakukan menggunakan algoritma SVM (*Support Vector Machine*) yang bekerja dalam tiga kernel, yaitu linear, polynomial dan rbf dengan akurasi yang didapatkan berturut-turut adalah 86,5%, 87,5% dan 89% [11]. Klasifikasi akun Twitter yang terindikasi spam juga pernah dilakukan menggunakan berbagai macam algoritma, di antaranya SVM, KNN, XGB, *Random Forest*, dan *Extra Trees*. Dari hasil penelitian ini, algoritma *Random Forest* memiliki akurasi tertinggi di nilai 99,30%, lalu diikuti oleh algoritma XGB dengan akurasi 98,39%, disusul oleh algoritma *Extra Trees* dan KNN dengan akurasi 98,02% dan algoritma SVM dengan nilai 95% [12].

Berdasarkan penelitian-penelitian dan penjelasan yang telah disampaikan, maka pada penelitian ini akan melakukan klasifikasi akun *buzzer* menggunakan algoritma *K-Nearest Neighbor* (KNN) pada tagar #STYTanpaDiasporaNol guna mendeteksi akun-akun provokator yang berinteraksi di media sosial X.

TINJAUAN PUSTAKA

X (Twitter)

X atau yang dulu dikenal sebagai Twitter adalah platform media sosial yang memungkinkan para penggunanya mengunggah pesan teks singkat, foto ataupun video, disebut *tweet*, untuk berbagi informasi dan opini. X memiliki fokus dalam pengembangan media sosial mereka pada pengalaman pengguna terhadap peristiwa terkini, seperti foto, video dan tautan yang dapat dibagikan dengan pengguna lainnya [13]. Penggunanya yang cukup banyak menjadikan X menjadi wadah untuk masyarakat dalam menyampaikan pandangan, opini maupun dukungan terhadap permasalahan tertentu, contohnya politik [14].

Buzzer

Buzzer merupakan istilah untuk akun yang menciptakan efek *buzz* (dengung) yang menyebabkan diskursus publik [15]. Dalam media sosial, *buzzer* adalah pengguna media sosial yang dapat mempengaruhi orang lain melalui pesan yang ia unggah [16]. Adapun karakteristik yang dimiliki oleh *buzzer* menurut [17] sebagai berikut:

1. *Buzzer* mengarah kepada perilaku berisik dan memperbanyak suara, yang bertujuan untuk berpartisipasi dalam *retweet*, *posting* dan komentar.
2. *Buzzer* umumnya anonim dan menggunakan foto profil menarik.
3. *Buzzer* umumnya hanya mengikuti arus yang ada.

Pembelajaran Mesin (Machine Learning)

Pembelajaran mesin atau *machine learning* merupakan bagian dari kecerdasan buatan (*artificial intelligence*) yang bertujuan untuk membangun sistem yang dapat belajar dan membuat keputusan. Menurut Goldberg dan Holland (1988), definisi *machine learning* adalah aplikasi komputer dan algoritma matematika yang dibentuk berdasarkan cara pembelajaran dari data dan akan menghasilkan prediksi di masa yang akan datang [18]. Dalam *machine learning*, sebuah program komputer diberikan serangkaian tugas untuk diselesaikan dan meningkatnya kinerja *machine learning* diukur berdasarkan tugas yang diberikan dari waktu ke waktu dengan semakin banyaknya latihan yang dilakukan [19].

K-Nearest Neighbor

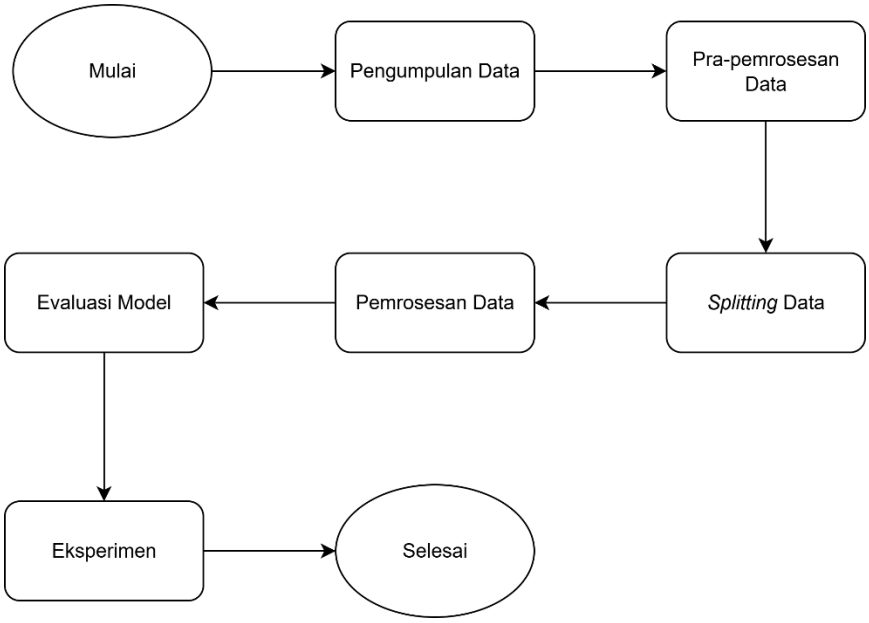
Algoritma KNN atau *K-Nearest Neighbor* adalah metode yang digunakan untuk melakukan klasifikasi terhadap suatu objek berdasarkan data pelatihan yang memiliki jarak terdekat dengan objek tersebut [20]. KNN bekerja dengan mencari K titik data yang terdekat dari suatu sampel yang diberikan, kemudian KNN menggunakan label kelas dari titik data tersebut untuk memprediksi label kelas dari sampel tersebut [21].

Dalam melakukan perhitungan jarak, algoritma KNN memiliki beberapa cara, salah satu di antaranya adalah *Euclidean Distance*. Adapun rumus *Euclidean Distance* sebagai berikut:

$$euc = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \quad (1)$$

Keterangan:
euc : Jarak *Euclidean* antara dua titik
p_i : Data *train* ke-i
q_i : Data *test* ke-i
i : Variabel data
n : Jumlah dimensi data

METODOLOGI



Gambar 1. Alur Proses Penelitian

Pengumpulan Data

Proses pengumpulan data pada penelitian ini dilakukan menggunakan teknik *scraping* pada media sosial X. Alat yang digunakan dalam *scraping* data adalah twscrape yang merupakan salah satu *library* dari Python. *Library* twscrape bekerja melalui API resmi sehingga diperlukan akun untuk melakukan *scraping tweet* di X. Data yang akan dikumpulkan dari media sosial X berupa informasi akun dari *user* yang mengunggah *tweet* dengan tagar #STYTanpaDiasporaNol. *Tweet* yang dikumpulkan mulai dari tanggal 21 Desember 2024 hingga tanggal 28 Desember 2024.

Pra-pemrosesan Data

Tahapan pra-pemrosesan data dilakukan untuk memilah informasi yang dibutuhkan dan menghapus informasi yang tidak diperlukan dalam proses klasifikasi. Pada tahap ini, *tweet* yang tidak berhubungan dan tidak sesuai topik dengan kasus *buzzer* pada tagar #STYTanpaDiasporaNol akan dihapus. Kemudian proses pelabelan akun dilakukan berdasarkan ciri-ciri atau karakteristik yang dimilik oleh akun. Label yang digunakan adalah *buzzer* dan *non-buzzer*. Berdasarkan karakteristik yang ada, sebuah akun mendapat label *buzzer* apabila telah memenuhi seluruh karakteristik dan sebaliknya untuk label *non-buzzer*.

Pada tahapan ini juga dilakukan rekayasa fitur (*feature engineering*) untuk meningkatkan akurasi yang dimiliki oleh model nantinya. Adapun rekayasa fitur yang dilakukan menghasilkan fitur-fitur baru seperti usia akun, frekuensi *tweet*, reputasi, rasio *follower* dan *friend*, tingkat pertumbuhan *follower*, tingkat pertumbuhan *friend*, dan rasio media dengan rumus persamaan berikut [10], [11], [12].

- Usia Akun

$$Waktu\ data\ diambil - Waktu\ akun\ dibuat \tag{2}$$

- Frekuensi *Tweet*

$$\frac{Total\ Tweet}{Usia\ Akun} \tag{3}$$

- Reputasi

$$\frac{Follower}{Friend+Follower} \tag{4}$$

- Rasio *Follower* dan *Friend*

$$\frac{Follower}{Friend} \tag{5}$$

- Tingkat Pertumbuhan *Follower*

$$\frac{Follower}{Usia\ Akun} \tag{6}$$

- Tingkat Pertumbuhan *Friend*

$$\frac{Friend}{Usia\ Akun} \tag{7}$$

- Rasio Media

$$\frac{Total\ Media}{Total\ Tweet} \tag{8}$$

Proses pra-pemrosesan data diakhiri dengan melakukan standarisasi pada seluruh data yang ada untuk mendapatkan akurasi model yang lebih baik karena telah mengatasi skala yang sangat besar atau sangat kecil.

Splitting Data

Kemudian pada tahapan *splitting* data, seluruh set data yang berupa informasi akun dari *tweet* pada tagar #STYTanpaDiasporaNol, akan dibagi menjadi dua set data yang berbeda yaitu data *train* dan data *test*. Perbandingan jumlah data *train* dengan data *test* akan dilakukan dengan eksperimen untuk mendapatkan akurasi yang terbaik. Eksperimen pembagian data *train* dan data *test* yang akan dilakukan adalah 90:10, 80:20 dan 70:30.

Pemrosesan Data

Pemrosesan data merupakan tahap di mana model akan dibentuk dan dibangun berdasarkan dari set data yang ada. Algoritma KNN yang digunakan dalam penelitian ini, mengimplementasikan metode *Euclidean Distance* untuk menghitung jarak antar data. Penentuan parameter K diperlukan untuk mendapatkan hasil yang maksimal pada model klasifikasi. Eksperimen dilakukan dalam menentukan parameter K, di antaranya 3, 5, 7 dan 9 yang akan digunakan pada pembangunan model *machine learning*.

Evaluasi Model

Setelah dilakukannya proses klasifikasi, evaluasi model diperlukan untuk mengetahui bagaimana performa dan kinerja model dalam melakukan klasifikasi pada data. Evaluasi model dilakukan dengan menggunakan beberapa metrik, seperti akurasi (*accuracy*), presisi (*precision*), sensitivitas (*recall*) dan *F1-Score*. *Accuracy* merupakan persentase prediksi yang benar dari total prediksi yang dilakukan model. *Precision* adalah persentase perbandingan prediksi positif yang benar terhadap seluruh prediksi positif yang ada. *Recall* adalah persentase perbandingan prediksi positif yang benar terhadap seluruh data yang seharusnya positif. *F1-Score* adalah persentase gabungan dari *precision* dan *recall*.

Eksperimen

Tahapan eksperimen bertujuan untuk mengatasi ketidakseimbangan pada set data. Untuk mengatasi ketidakseimbangan tersebut, maka pengurangan (*undersampling*) dan penambahan (*oversampling*) data akan dilakukan. *Undersampling* data akan dilakukan terhadap data mayoritas yaitu kategori *buzzer* sedangkan *oversampling* data akan dilakukan terhadap data minoritas yaitu kategori *non-buzzer*. Proses *undersampling* dan *oversampling* akan dilakukan menggunakan *library* dari Python yaitu imblearn yang menangani ketidakseimbangan data. Untuk *undersampling* akan menggunakan metode TomekLinks sedangkan untuk *oversampling* akan menggunakan metode SMOTE.

HASIL DAN PEMBAHASAN

Pengumpulan Data

Penelitian yang dilakukan menggunakan data dari sosial media X yang berupa informasi akun pada setiap akun yang berinteraksi atau terlibat dalam *tweet* menggunakan tagar #STYTanpaDiasporaNol dengan rentang tanggal 21 Desember hingga 28 Desember 2024. Pengumpulan data dilakukan dengan teknik *scraping* data menggunakan salah satu *library* dari bahasa Python yaitu twscrape. Data yang didapatkan dari hasil proses *scraping* sebanyak 680 *tweet* pada rentang 7 hari.

Tabel 1. Sampel Data *Tweet* Pada Tagar

No.	*tweet_content*	*username*	...	media
1	Bahaya dg pendukung sty iQ 78...	tengsaw1	...	184
2	Oh katanya ada *buzzer*² yg diduga dari...	ynjaya	...	841
3	Duh, hmm… 🤔 🧐 \n\n📸:...	labiebsadat	...	1,960
4	@gilabola_ina #STYTanpaDiasporaNol	bangpoerslalu	...	7,948
5	⚽ 🕊️ Jagat media sosial Indonesia...	joedwip	...	37
...	...	...	...	...
676	Fans kecewa, bukan cuma karena kalah...	daccaaf	...	43
677	ngapa deh si coach eksperimen terus...	laginuscha	...	74
678	itu biar apa deh coach? Biar kita makin...	Delusilusii	...	86
679	@salmaindriana_ hadehhh dah pokoknya...	gallsxx_	...	321
680	@sebutir_baja emang lawak baanget yee...	gallsxx_	...	321

Pra-pemrosesan Data

Tahapan ini diperlukan untuk melakukan pengolahan informasi data yang dapat digunakan dalam model *machine learning*. Terdapat 4 proses penting yang akan dilakukan dalam tahapan pra-pemrosesan data, di antaranya yaitu *cleaning* data, pelabelan, rekayasa fitur, dan standarisasi.

Pembersihan atau *cleaning* data diperlukan untuk menghapus duplikasi data serta membuang informasi yang tidak diperlukan dan tidak sesuai dengan spesifikasi yang dibutuhkan. Pada proses pembersihan data, *tweet* akan dilakukan

 https://doi.org/10.56211/blendsains.v4i2.1093

filtering dengan membuang duplikat akun yang melakukan *tweet* lebih dari satu kali. Kemudian *tweet* yang disimpan merupakan *tweet* yang pertama kali diunggah oleh akun, sehingga didapatkan akun dengan *tweet* yang bersifat unik.

Tahapan selanjutnya adalah melakukan *filtering* terhadap *tweet* yang tidak sesuai dengan topik dengan tujuan untuk menghapus *outlier* dari data yang tidak sesuai dengan konteks sepakbola Indonesia. Sehingga hasil yang dapat digunakan dalam data untuk pembangunan model yang bersifat unik dan sesuai konteks.

Hasil proses *cleaning* data didapatkan sebanyak 573 *tweet* yang mewakili tiap akun serta sesuai dengan topik yang ada.

Tabel 2. *Tweet* Setelah Proses Cleaning Data

No.	*tweet_content*	*username*	...	media
1	Bahaya dg pendukung sty iQ 78...	tengsaw1	...	184
2	Oh katanya ada *buzzer*² yg diduga dari...	ynjaya	...	841
3	@gilabola_ina #STYTanpaDiasporaNol	bangpoerslalu	...	7948
4	#styout #STYTanpaDiasporaNol Kalah...	pram35708	...	0
5	Tagar #ErickOut tiba-tiba muncul...	skorindonesia	...	20129
...	...	...	...	...
569	@sebutir_baja kok lawak malah jadi...	piipels	...	259
570	Fans kecewa, bukan cuma karena kalah...	daccaaf	...	43
571	ngapa deh si coach eksperimen terus...	laginuscha	...	74
572	itu biar apa deh coach? Biar kita makin...	Delusilusii	...	86
573	@sebutir_baja emang lawak baanget yee...	gallsxx_	...	321

Tahap pelabelan dilakukan secara manual berdasarkan dengan spesifikasi yang dimiliki oleh akun. Spesifikasi yang digunakan sebagai berikut.

1. Akun *buzzer* umumnya anonim, bukan merupakan nama asli atau organisasi.
2. Akun *buzzer* biasanya merupakan akun baru, usia akun di bawah 3 tahun.
3. Pola dan frekuensi unggahan pada akun *buzzer* terstruktur dengan tagar relevan.
4. Akun *buzzer* melakukan *retweet* atau *reply* pada satu topik unggahan umumnya produk disertakan dengan tagar khusus.

Tabel 3. Akun Setelah Pelabelan

No.	*username*	...	label
1	tengsaw1	...	*non-buzzer*
2	ynjaya	...	*non-buzzer*
3	bangpoerslalu	...	*non-buzzer*
4	pram35708	...	*non-buzzer*
5	skorindonesia	...	*non-buzzer*
...	...	...	...
569	piipels	...	*buzzer*
570	daccaaf	...	*buzzer*
571	laginuscha	...	*buzzer*
572	Delusilusii	...	*buzzer*
573	gallsxx_	...	*buzzer*

Rekayasa Fitur merupakan sebuah tahapan untuk menciptakan fitur-fitur baru dengan cara mengolah fitur-fitur yang talah ada dengan tujuan meningkatkan akurasi pada model. Adapun berikut merupakan seluruh rekayasa fitur yang dilakukan pada setiap 573 data yang telah dikomputasikan.

Tabel 4. Data Setelah Proses Rekayasa Fitur

No.	*Followers*	...	Usia Akun	Frekuensi *Tweet*	Reputasi	...	label
1	59	...	4561	1.9520	0.4876	...	*non-buzzer*
2	544	...	5551	13.1130	0.6348	...	*non-buzzer*
3	2240	...	4352	11.2307	0.5386	...	*non-buzzer*
4	0	...	105	0.0857	0	...	*non-buzzer*
5	29095	...	4571	38.7204	0.9896	...	*non-buzzer*
...	...	...	...	...	...	...	...
569	1015	...	872	2.7144	0.4810	...	*buzzer*

570	1170	...	558	0.8889	0.4924	...	*buzzer*
571	1276	...	636	1.7972	0.8812	...	*buzzer*
572	1303	...	680	2.1397	0.8038	...	*buzzer*
573	9273	...	4465	4.4287	0.8996	...	*buzzer*

Standarisasi merupakan tahap terakhir dari pra-pemrosesan data. Standarisasi pada data bertujuan untuk meningkatkan akurasi serta kecepatan dalam melakukan pelatihan model dikarenakan data yang sangat besar atau sangat kecil akan memiliki rentang yang cukup kecil setalah dilakukannya standarisasi. Berikut merupakan hasil data setelah dilakukan standarisasi *z-score*.

Tabel 5. Akun Setelah Standarisasi

No.	*Followers*	Friends	Statuses	Media	Usia Akun	...	label
1	-0.0495	-0.3441	0.3014	-0.0204	2.6369	...	*non-buzzer*
2	-0.0427	0.1373	4.1735	0.6047	3.3703	...	*non-buzzer*
3	-0.0190	3.2172	2.7241	7.3676	2.4820	...	*non-buzzer*
4	-0.0503	-0.4572	-0.2376	-0.1955	-0.6646	...	*non-buzzer*
5	0.3563	0.1239	10.4888	18.9588	2.6443	...	*non-buzzer*
...	...	...	...	...	...	...	...
569	-0.0361	1.6370	-0.0947	0.0509	-0.0963	...	*buzzer*
570	-0.0339	1.8499	-0.2081	-0.1546	-0.3289	...	*buzzer*
571	-0.0325	-0.1331	-0.1689	-0.1251	-0.2711	...	*buzzer*
572	-0.0321	0.1469	-0.1500	-0.1137	-0.2385	...	*buzzer*
573	0.0793	1.5219	0.9603	0.1099	2.5657	...	*buzzer*

Splitting Data

Penelitian ini menggunakan 3 macam rasio pembagian data yang akan dilaksanakan dalam eksperimen, di antaranya 70:30, 80:20, 90:10. Eksperimen ini dilakukan bertujuan untuk mendapatkan pembagian data dengan hasil yang paling baik dalam pembangunan model.

Tabel 6. *Splitting* Data Model

Jenis	70:30	80:20	90:10
Data *Train*	401	458	515
Data *Test*	172	115	58

Setelah proses *splitting* data dengan 3 varian rasio yang berbeda, maka didapatkan perbandingan kategori sebagai berikut.

Tabel 7. Perbandingan Antar Kategori

Kategori	70:30	80:20	90:10
Buzzer	314	356	399
Non-buzzer	87	102	116

Pemrosesan Data

Pada tahapan pemrosesan data, pembangunan model *machine learning* dilakukan untuk mendapatkan hasil akhir. Algoritma KNN bekerja dengan menghitung jarak terdekat terhadap K, dengan K merupakan jumlah objek terdekat. Penelitian ini menggunakan metode *Euclidean Distance* untuk menghitung jarak dalam algoritma KNN.

Proses pembangunan model dilakukan setelah simulasi perhitungan sampel data selesai. Dalam pembangunan model menggunakan Python, K ditentukan berdasarkan eksperimen dengan menggunakan 3, 5, 7 dan 9.

Tabel 8. Nilai Akurasi Seluruh Varian Model

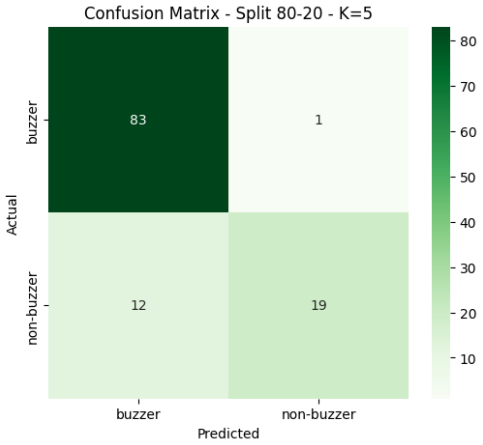
Nilai *K*	70:30	80:20	90:10
***K* = 3**	84%	86%	84%
***K* = 5**	86%	**89%**	84%
***K* = 7**	85%	86%	84%
***K* = 9**	85%	87%	86%

 https://doi.org/10.56211/blendsains.v4i2.1093

Dari keseluruhan hasil eksperimen pembangunan model, ditemukan model yang memiliki akurasi terbaik pada nilai 89% yaitu pada *splitting* data 80:20 dengan K = 5.

Evaluasi Model

Model yang telah dibangun akan dievaluasi untuk melihat kinerja dan kelayakan model agar dapat digunakan. Salah satu alat evaluasi model yang umum digunakan adalah *confusion matrix*. *Confusion Matrix* bekerja dengan menghitung nilai dari *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN).



Gambar 2. *Confusion Matrix* Model

Berdasarkan visualisasi *confusion matrix* model di atas, didapatkan jumlah TP sebanyak 83, FP sebanyak 1, FN sebanyak 12 dan TN sebanyak 19. Hal ini menandakan bahwa model berhasil memprediksi akun dengan benar sebanyak 102 akun dan salah memprediksi akun sebanyak 13 akun dari total 115 akun.

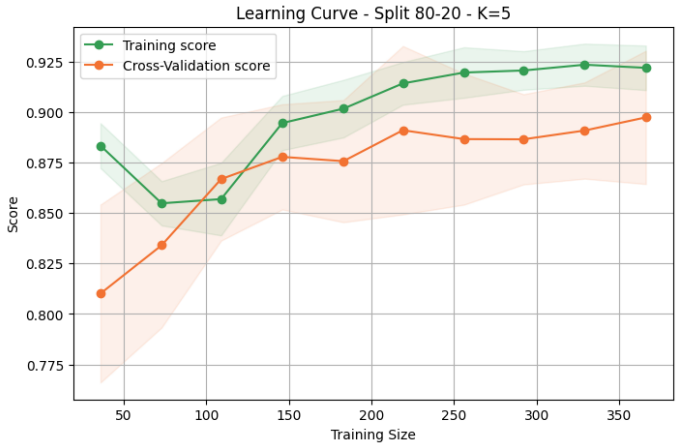
Kemudian evaluasi model selanjutnya melalui laporan klasifikasi atau *classification report* berisikan informasi berupa nilai dari *accuracy*, *precision*, *recall*, dan *F1-score* dalam model. Berdasarkan laporan yang ada ditemukan nilai sebagai berikut.

Tabel 9. *Classification Report* dari Model

Jenis	Evaluasi	Nilai
Buzzer	*Precision*	87%
	Recall	99%
	F1-score	93%
Non-buzzer	*Precision*	95%
	Recall	61%
	F1-score	75%
Macro Average	*Precision*	91%
	Recall	80%
	F1-score	84%
Weighted Average	*Precision*	89%
	Recall	89%
	F1-score	88%
Accuracy		89%

Berdasarkan dari tabel di atas, model sudah bekerja cukup baik dapat dilihat dari metrik evaluasi pada *macro* dan *weighted average* yang dimiliki model berada di rentang 84% hingga 91%. Namun terdapat ketimpangan nilai dalam mendeteksi masing-masing kategori, dapat dilihat bahwa *recall* yang dimiliki oleh kategori *buzzer* sebesar 99% sedangkan pada kategori *non-buzzer* hanya sebesar 61%. Model sangat baik dan hampir sempurna dalam memprediksi kategori *buzzer* berbanding terbalik dalam kategori *non-buzzer*, model cukup kesulitan dalam memprediksinya.

Metode evaluasi menggunakan *learning curve* bertujuan untuk menilai kelayakan model yang dibangun serta dapat melihat indikasi *overfitting* atau *underfitting* pada model.



Gambar 3. *Learning Curve* Model

Berdasarkan kurva di atas, *training* score pada model dimulai di atas angka 88%, walaupun sempat mengalami penurunan, kurva mengalami kenaikan dan mendapatkan nilai akhir di atas 92%. Sedangkan *cross-validation score*, nilai sudah berada di atas angka 80% sejak awal yang dapat mengindikasikan bahwa model terlalu cepat untuk memprediksi dengan tepat.

Eksperimen

Tahapan eksperimen bertujuan untuk mengatasi indikasi *overfitting* yang terjadi pada model yang telah dibangun. Indikasi *overfitting* ini terjadi disebabkan karena adanya ketidakseimbangan pada set data. Proses *undersampling* dan *oversampling* akan dilakukan menggunakan *library* dari Python yaitu imblearn yang menangani ketidakseimbangan data. Untuk *undersampling* akan menggunakan metode TomekLinks sedangkan untuk *oversampling* akan menggunakan metode SMOTE.

Tabel 10. Data Sebelum *Oversampling* dan *Undersampling*

Kategori	70:30	80:20	90:10
Buzzer	314	356	399
Non-buzzer	87	102	116

Tabel 11. Data Setelah *Oversampling* dan *Undersampling*

Kategori	70:30	80:20	90:10
Buzzer	307	348	389
Non-buzzer	307	348	389

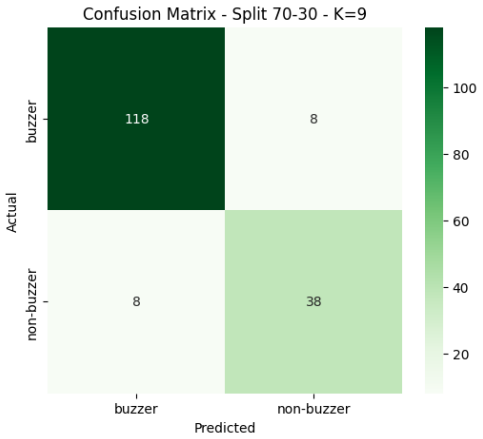
Proses pembangunan model *machine learning* pada Python dilakukan menggunakan data yang telah diseimbangkan dengan tetap menggunakan eksperimen dengan nilai K sebesar 3, 5, 7 dan 9.

Tabel 12. Nilai Akurasi Seluruh Varian Model Eksperimen

Nilai K	70:30	80:20	90:10
***K* = 3**	86%	83%	84%
***K* = 5**	89%	87%	86%
***K* = 7**	88%	89%	86%
***K* = 9**	**91%**	88%	86%

Hasil yang didapatkan adalah varian dengan *splitting* data 70:30 dengan K = 9 mendapatkan akurasi yang terbaik sebesar 91%.

 https://doi.org/10.56211/blendsains.v4i2.1093



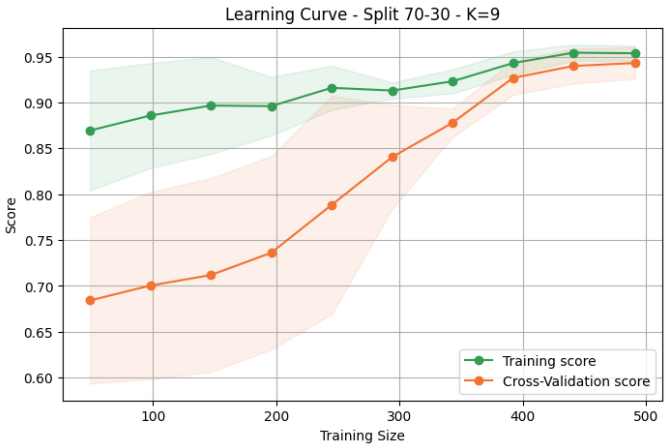
Gambar 4. *Confusion Matrix* Model Eksperimen

Confusion Matrix pada gambar menunjukkan bahwa terdapat jumlah TP sebanyak 118, FP sebanyak 8, FN sebanyak 8 dan TN sebanyak 38. Model ini dapat memprediksi akun dengan benar sebanyak 156 akun dan salah memprediksi akun sebanyak 16 akun.

Tabel 13. *Classification Report* dari Model Eksperimen

Jenis	Evaluasi	Nilai
Buzzer	*Precision*	94%
	Recall	94%
	F1-score	94%
Non-buzzer	*Precision*	83%
	Recall	83%
	F1-score	83%
Macro Average	*Precision*	88%
	Recall	88%
	F1-score	88%
Weighted Average	*Precision*	91%
	Recall	91%
	F1-score	91%
Accuracy		91%

Berdasarkan dari tabel di atas, dapat dilihat bahwa terdapat peningkatan kestabilan pada keseluruhan hasil evaluasi metrik. Peningkatan *recall* pada kategori minoritas yaitu *non-buzzer* juga mengalami peningkatan pada angka 83%. Berdasarkan hasil dari *classification report* ini, tidak ditemukan indikasi *overfitting* seperti pada model sebelumnya karena model sudah cuku baik dalam memprediksi kedua kategori tersebut.



Gambar 5. *Learning Curve* Model Eksperimen

Adapun dari hasil *learning curve* yang ada, ditemukan bahwa *cross-validation score* dimulai dari angka yang lebih realistis yaitu dimulai pada angka di bawah 70% yang seiring jumlah latihan meningkat hingga hampir menyentuh 95%. Adapun jarak antar *training* score dengan *cross-validation score* sudah cukup dekat yang menandakan model sudah berjalan dengan cukup baik.

Nilai akurasi model eksperimen KNN terbaik mendapatkan nilai sebesar 91%, akurasi ini lebih tinggi dibandingkan dengan model *Decision Tree* yang mendapatkan akurasi dengan nilai 88,5% [10] serta model SVM yang dilakukan pada tiga kernel berbeda dengan akurasi 86,5%; 87,5% dan 89% [11].

KESIMPULAN DAN SARAN

Berdasarkan hasil dari pembangunan model *machine learning* menggunakan algoritma KNN dengan metode eksperimen mendapatkan varian pada *split* data 80:20 dengan K = 5 dan nilai akurasi yang didapatkan sebesar 89% dengan *precision*, *recall* dan *accuracy* sebesar 89%. Namun terdapat kekurangan model dalam mendeteksi akun, model sangat baik dalam memprediksi kelas mayoritas namun kesulitan dalam memprediksi kelas minoritas. Hal ini menunjukkan adanya indikasi overfitting yang disebabkan ketidakseimbangan data. Eksperimen dilakukan untuk mendapatkan hasil yang lebih baik pada model dengan melakukan *undersampling* dan *oversampling* pada data yang bertujuan membuat keseimbangan data. Hasil yang didapatkan bahwa varian pada *split* data 70:30 dengan K = 9 diperoleh akurasi sebesar 91% dengan *precision*, *recall* dan *accuracy* juga sebesar 91%. Model eksperimen ini juga cukup baik mendeteksi kelas mayoritas maupun kelas minoritas. Saran untuk penelitian selanjutnya adalah mengembangkan penelitian ini menggunakan algoritma dengan tingkat kompleksitas yang lebih tinggi untuk memperoleh hasil yang lebih baik, serta melakukan pemilihan dan rekayasa fitur yang lebih tepat untuk meningkatkan kecerdasan model dalam menangani variansi objek.

DAFTAR PUSTAKA

[1] We Are Social dan Kepios, “Digital 2024: Indonesia,” We Are Social. Diakses: 13 Januari 2025. [Daring]. Tersedia pada: https://wearesocial.com/id/blog/2024/01/digital-2024/

[2] W. Aulia dan S. Lexianingrum, “Analisis Fenomena Buzzer pada Konten Media Sosial Tiktok Menjelang Pemilu 2024,” *IJM: Indonesian Journal of Multidisciplinary*, vol. 2, no. 4, Jun 2024, Diakses: 31 Januari 2025. [Daring]. Tersedia pada: https://journal.csspublishing.com/index.php/ijm/article/view/782

[3] A. Faulina, E. Chatra, dan Sarmiati, “Peran Buzzer dalam Proses Pembentukan Opini Publik di New Media,” *Jurnal Pendidikan Tambusai*, vol. 5, no. 2, hlm. 2806–2820, Jul 2021, Diakses: 13 Januari 2025. [Daring]. Tersedia pada: https://jptam.org/index.php/jptam/article/view/1305

[4] N. R. Yunus, I. Susilowati, dan Z. Zahrotunnimah, “Kecebong Versus Kampret; Slogan Negatif Dalam Komunikasi Politik Pada Pemilihan Presiden 2019,” *SALAM: Jurnal Sosial dan Budaya Syar-i*, vol. 6, no. 4, hlm. 404–416, Des 2019, doi: 10.15408/sjsbs.v6i4.13747.

[5] N. Khotimah, “Geger Tagar ‘STY Tanpa Diaspora Nol’ Diduga Ulah Buzzer, Sekali Posting Dapat Bayaran Menggiurkan?,” *Suara*, 23 Desember 2024. Diakses: 31 Januari 2025. [Daring]. Tersedia pada: https://www.suara.com/lifestyle/2024/12/23/154757/geger-tagar-sty-tanpa-diaspora-nol-diduga-ulah-buzzer-sekali-posting-dapat-bayaran-menggiurkan

[6] D. Purnomo, “Viral Tagar ‘STY Tanpa Diaspora Nol’ Menggema di Sosial Media, Diduga Ulah Buzzer yang Ingin Serang Shin Tae-yong,” *Solo Balapan*, 23 Desember 2024. Diakses: 31 Januari 2025. [Daring]. Tersedia pada: https://solobalapan.jawapos.com/sport/2305452564/viral-tagar-sty-tanpa-diaspora-nol-menggema-di-sosial-media-diduga-ulah-buzzer-yang-ingin-serang-shin-tae-yong

[7] A. Fauzi, “Perang Tagar #STYOut dan #STYStay Ramai di Medsos atas Hasil Piala AFF 2024, STY Tetap Stay atau Good Bye?,” *Media Indonesia*, 25 Desember 2024. Diakses: 31 Januari 2025. [Daring]. Tersedia pada: https://mediaindonesia.com/sepak-bola/729104/perang-tagar-styout-dan-stystay-ramai-di-medsos-atas-hasil-piala-aff-2024-sty-tetap-stay-atau-good-bye

[8] A. Diahwahyuningtyas dan A. N. Dzulfaroh, “Pemecatan Shin Tae-yong Tuai Sorotan, Apa yang Perlu Diketahui?,” *Kompas*, 7 Januari 2025. Diakses: 31 Januari 2025. [Daring]. Tersedia pada: https://www.kompas.com/tren/read/2025/01/07/050000965/pemecatan-shin-tae-yong-tuai-sorotan-apa-yang-perlu-diketahui-

[9] F. Janati dan J. Carina, “DPR: Pemecatan Shin Tae-yong Harus Transparan dan Obyektif,” *Kompas*, Jakarta, 7 Januari 2025. Diakses: 31 Januari 2025. [Daring]. Tersedia pada:

https://nasional.kompas.com/read/2025/01/07/07373051/dpr-pemecatan-shin-tae-yong-harus-transparan-dan-obyektif

[10] A. Maulana dan S. Kuswayati, "Klasifikasi Akun Buzzer Pemilu Pada Media Sosial Twitter Berdasarkan Data Tweet Menggunakan Algoritma C4.5," *Naratif Jurnal Nasional Riset Aplikasi dan Teknik Informatika*, vol. 3, no. 02, hlm. 30–35, Des 2021, doi: 10.53580/naratif.v3i02.132.

[11] A. Mustofa *dkk.*, "Twitter Buzzer Detection System Using Tweet Similarity Feature and Support Vector Machine," *NJCA (Nusantara Journal of Computers and Its Applications)*, vol. 8, no. 1, hlm. 7, Jun 2023, doi: 10.36564/njca.v8i1.306.

[12] S. J. Alsunaidi, R. T. Alraddadi, dan H. Aljamaan, "Twitter Spam Accounts Detection Using Machine Learning Models," dalam *2022 14th International Conference on Computational Intelligence and Communication Networks (CICN)*, IEEE, Des 2022, hlm. 525–531. doi: 10.1109/CICN56167.2022.10008339.

[13] M. Kholilullah, M. Martanto, dan U. Hayati, "Analisis Sentimen Pengguna Twitter (X) Tentang Piala Dunia Usia 17 Menggunakan Metode Naive Bayes," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, hlm. 392–398, Feb 2024, doi: 10.36040/jati.v8i1.8378.

[14] I. L. Achmadi, H. Yuana, dan M. F. Rahmat, "Analisis Sentimen Hasil Pemilihan Presiden Ri 2024 Pada Twitter Atau X Dengan Metode Naive Bayes Classifier," *INNOVATIVE: Journal Of Social Science Research*, vol. 4, no. 5, hlm. 9033–9048, Okt 2024, doi: https://doi.org/10.31004/innovative.v4i5.15571.

[15] S. Sisrilnardi dan M. Nur Alamsyah, "Peran Buzzer Sebagai Opinion Makers Dalam Proses Reklamasi Teluk Jakarta Tahun 2016-2017," *SIBATIK JOURNAL: Jurnal Ilmiah Bidang Sosial, Ekonomi, Budaya, Teknologi, dan Pendidikan*, vol. 2, no. 3, hlm. 839–854, Feb 2023, doi: 10.54443/sibatik.v2i3.670.

[16] C. Juditha, "Buzzer di Media Sosial: Antara Marketing Politik dan Black Campaign dalam Pilkada," dalam *New Media & Komunikasi Politik (Telaah Kontestasi Politik dalam Ruang New Media)*, 1 ed., D. H. Santoso, Ed., Yogyakarta: Mbridge Press, 2018, 1, hlm. 1–24.

[17] M. Rizki, Y. Fajar Wulandari, dan S. Himawan, "Analisis Peran Buzzer Dalam Membentuk Citra Merek/Produk Di Media Sosial Instagram," *INNOVATIVE: Journal Of Social Science Research*, vol. 4, no. 5, hlm. 6530–6540, Okt 2024, doi: https://doi.org/10.31004/innovative.v4i5.15686.

[18] A. Roihan, P. A. Sunarya, dan A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper," *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 5, no. 1, Mei 2020, doi: 10.31294/ijcit.v5i1.7951.

[19] M. M. Taye, "Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions," *Computers*, vol. 12, no. 5, hlm. 91, Apr 2023, doi: 10.3390/computers12050091.

[20] R. Harun, K. Chandra Pelangi, dan Y. Lasena, "Penerapan Data Mining Untuk Menentukan Potensi Hujan Harian Dengan Menggunakan Algoritma K Nearest Neighbor (KNN)," *Jurnal Manajemen informatika & Sistem Informasi)*, vol. 3, no. 1, hlm. 2614–1701, Jan 2020, doi: https://doi.org/10.36595/misi.v3i1.125.

[21] S. D. Prasetyo, S. S. Hilabi, dan F. Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *Jurnal KomtekInfo*, hlm. 1–7, Jan 2023, doi: 10.35134/komtekinfo.v10i1.330.