Indonesian Journal of Data and Science (IJODAS) ISSN: 2715-9930
Vol 3, No 2, July 2022, pp. 61-71

Analisis sentimen terhadap *body shaming* pada twitter menggunakan metode *Naïve Bayes Classifier*

St. Fajriah Fattah[a,1], Purnawansyah[a,2]

[a] *Universitas Muslim Indonesia, Jl. Urip Sumoharjo KM.05, Makassar dan 90231, Indonesia*
[1] *13020170238@umi.ac.id;* [2] *purnawansyah@umi.ac.id;*

INFORMASI ARTIKEL

Diterima : 15 – 05 – 2022
Direvisi : 19 – 06 – 2022
Diterbitkan : 31 – 07 – 2022

Kata Kunci:
Twitter
Naïve Bayes Classifier
Performansi
Klasifikasi
Data Mining

ABSTRAK

Salah satu bentuk media sosial yang sedang populer saat ini adalah *twitter*. Namun tidak jarang pengguna *twitter* memberikan komentar yang cenderung menyinggung pengguna *twitter* lain dengan kalimat negatif. Salah satu bentuk komentar negatif yang sering dilontarkan pengguna *twitter* adalah tentang *body shaming*. *Body shaming* merupakan komentar negatif terhadap fisik seseorang seperti gendut, pesek, cungkring dan lain-lain. Berdasarkan perilaku *body shaming* pada *twitter*, maka pada penelitian ini akan dilakukan analisis sentimen menggunakan metode *Naive Bayes Classifier*. Tujuan dari penelitian adalah mengukur performa *Accuracy, Precision, Recall*, dan *f-measure* pada metode *Naïve bayes classifier* dalam analisis sentimen terhadap *body shaming* pada *Twitter*. Dataset tersebut digunakan untuk mengklasifikasikan *tweet s* yang bersifat positif dan negatif. Teknik klasifikasi yang digunakan yaitu dengan mengukur performa dari *Accuracy, Precision, Recall*, dan *f-measure* menggunakan metode *naïve bayes classifier*. Berdasarkan hasil pengujian performansi *Accuracy, Precision, Recall*, dan *f-measure* dengan *feature model trigram* menggunakan metode *naïve bayes classifier* dilakukan pada dataset *tweet s body shaming* yang berjumlah 908 data. Berdasarkan hasil pengujian performa dengan model trigram didapatkan hasil *Accuracy* 61%, *Precision* 56%, *Recall* 55% dan *f-measure* 55%.

I. Pendahuluan

Perkembangan teknologi saat ini membuat semua informasi semakin mudah untuk diakses dan telah merubah kecenderungan kebiasaan seseorang dalam mengekspresikan opininya [1]. Penggunaan media sosial telah membawa pengaruh dalam perubahan kebudayaan, gaya hidup, bahasa serta pergaulan masyarakat sehari-hari. *Twitter* adalah salah satu media sosial hasil pemikiran Dorsey ini yang sangat diminati saat ini. Negara Indonesia diklaim menjadi salah satu negara yang pertumbuhan pengguna aktif harian *twitter*-nya paling besar berdasarkan laporan finansial *twitter* kuartal ke-3 tahun 2019[2]. Masyarakat Indonesia banyak menggunakan *twitter* dalam seperti berbagi informasi pemberitaan, ajang promosi dan bisnis atau hanya sekedar tempat memberikan *tweet* , komentar, fakta ataupun opini[3]. Namun tidak jarang bahwa pengguna *Twitter* memberikan *tweet*, komentar, fakta ataupun opini yang tidak mengenakan dan cenderung menyinggung pengguna *Twitter* lain dengan kata-kata atau kalimat yang mengandung ke arah yang negatif. Salah satu bentuk *tweet* atau komentar negatif yang sering digunakan pengguna *Twitter* adalah *Body Shaming*[4].

Body Shaming atau mengomentari kekurangan fisik orang lain tanpa disadari sering dilakukan orang-orang.[5] *Body shaming* sudah termasuk jenis perundungan secara verbal atau lewat kata-kata dan termasuk sebagai tindakan perundungan yang terkait dengan tampilan fisik seseorang atau lebih[6]. Contoh *body shaming* adalah penyebutan dengan gendut, pesek, cungkring, dan lain sebagainya yang berkaitan dengan tampilan fisik. Terdapat 966 kasus penghinaan fisik atau *body shaming* yang ditangani polisi dari seluruh Indonesia sepanjang 2018. Sebanyak 347 kasus di antaranya selesai, baik melalui penegakan hukum maupun pendekatan mediasi antara korban dan pelaku. Alasan pengguna *Twitter* melakukan *body shaming* pun beragam seperti sebagai bahan canda tawa, mencairkan suasana, sekadar iseng atau ikut-ikutan pengguna *Twitter* lain, hingga memang ingin menjelek-jelekan seseorang[7]. Dampak dari *Body Shaming* yaitu menurunkan rasa percaya diri, menimbulkan gangguan depresi, dan yang lebih fatalnya dapat meningkatkan risiko bunuh diri pada korban *body Shaming*[8]. Bentuk analisis yang dapat dilakukan untuk menganalisis

perilaku pengguna *Twitter* adalah analisis sentimen. Analisis sentimen merupakan suatu proses yang dilakukan untuk memahami, mengekstrak, dan mengolah data berupa tekstual secara otomatis yang digunakan untuk mendapatkan informasi[9]. Salah satunya dengan menganalisis sentimen pengguna *Twitter* dengan menggunakan hasil *tweet* s dari pengguna *Twitter* yang melakukan *Body Shaming*. Pada berbagai penelitian tentang analisis dokumen tekstual yang sudah pernah dilakukan oleh Taheri dan Mammodov (2013), Hall (2007), dan Shahheidari dkk (2013) paling banyak menggunakan metode *Naïve Bayes Classifier*.

Naïve bayes classifier adalah salah satu algoritma yang popular digunakan untuk keperluan *data mining* karena kemudahan penggunaannya[10]. Pada klasifikasi Naïve Bayes, proses pembelajaran lebih ditekankan pada mengestimasi probabilitas[11]. Serta waktu pemrosesan yang cepat, mudah diimplementasikan dengan strukturnya yang cukup sederhana dan tingkat efektifitas yang tinggi. *Naïve Bayes* adalah metode klasifikasi dalam penambangan teks yang dapat digunakan juga dalam analisis sentiment dan hanya membutuhkan jumlah data pelatihan (training data) yang kecil[12]. Sentiment analysis merupakan sebuah metode yang digunakan untuk memahami, mengekstrak data opini, dan mengolah data tekstual secara otomatis untuk mendapatkan sebuah sentiment yang terkandung dalam sebuah opini[13], [14]. Pada penelitian yang dilakukan[15] yang berjudul "Analisis Sentimen dari *Body Shaming Beauty Vlog* Komentar" bahwa perbandingan data latih dan data uji 90:10. Skenario ini mencapai akurasi 98,48%, *Recall* 99,53%, presisi 98,90% dan f- *measure* 99,21%. Namun saat ini belum ada penelitian mengenai sentimen analisis *body shaming* pada *twitter* Bahasa Indonesia menggunakan *Naïve Bayes Classifier*.

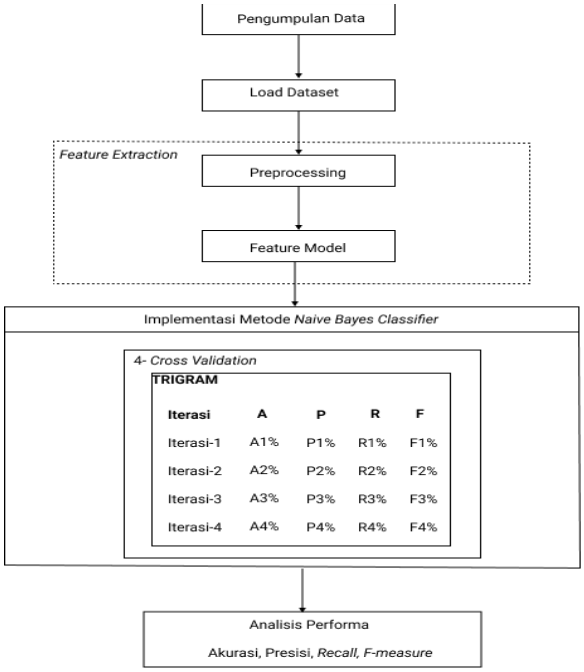
Berdasarkan masalah yang diuraikan sebelumnya, maka pada penelitian ini akan melakukan analisis sentimen terhadap *body shaming* pada *twitter* menggunakan *Naïve Bayes Classifier*. Dengan metode evaluasi yang akan dilakukan yaitu mengukur performa metode *Bayes* menggunakan *confusion matriks* untuk memperoleh nilai akurasi, presisi, *Recall*, dan *f-measure*. Penelitian ini membatasi penelitian dengan hanya menggunakan metode klasifikasi *Naïve Bayes Clasifier* dengan menggunakan dataset *tweet s* dalam Bahasa Indonesia pada *social media twitter* dengan bantuan *Twitter* API, *Feature* model yang digunakan ialah tiagram, dari beberapa data yang ambil 90% digunakan sebagai *data training* dan 10% sebagai *data testing*

II. Metode

Metode berisi tahapan atau prosedur penelitian dan algoritma yang digunakan dalam penelitian, berikut ini akan disajikan beberapa hal terkait dengan tahapan dalam metode penelitian:

A. Tahapan Penelitian

Tahapan dalam penelitian ini terdiri dari pengumpulan data *tweet* dari *twitter* dengan karakteristik data *body shaming* seperti cungkring, botak, tepos, pesek, dan lain sebagainya. Kemudian tahap pembacaan *dataset* (*load dataset*) yan mengandung kata-kata *body shiming*, ditemukan sebanyak 908 data, tahap berikutnya ialah *preprocessing* dengan tahapan *cleaning*, *case folding*, *tokenizing*, *stemming*, dan *stopword*. Tahap berikutnya ialah *feature model* (proses ekstrak data untuk menghasilkan fitur pada setiap *tweet s* yang akan digunakan pada tahap klasifikasi analisis sentimen. Tahap berikutnya ialah *cross validation* (tahap ini dataset akan dibagi menjadi dua bagian *data training* dan *data testing*, untuk *data training* 75% dari total dataset sedangkan *data testing* 25%. Tahap berikutnya ialah implementasi dari metode *naïve bayes lassifier* (tahap ini dilakukan pembobotan dari penggunaan metode dengan memanfaatkan pemrograman dengan Bahasa *phyton*. Tahap berikutnya analisis performa metode *naïve bayes classifier* (akurasi, presisi, *Recall*, dan *f-measure*), tahap terakhir ialah pengambilan keputusan terhadap hasil perhitungan pada tahap performa. Tahapan penelitian disajikan dalam bentuk bagan pada Gambar 1, berikut ini :



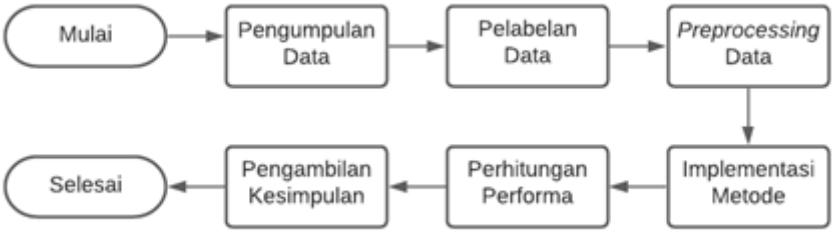
Gambar 1. Tahapan Penelitian

B. Cara Pengumpulan Data

Pengumpulan data perlu dilakukan terlebih dahulu, umumnya data yang diperoleh bersifat *big data* yang berarti pengumpulan data dari berbagai macam sumber yang relevan. Pengumpulan data pada penelitian ini menggunakan situs internet yaitu dari situs https://www.kaggle.com/ kaggle.

C. Teknik Analisis Data

Pada penelitian ini, metode yang akan digunakan berdasarkan jenis data yang di olah yaitu dengan menggunakan metode kuantitatif. Teknik analisis data ditunjukkan pada Gambar 2.



Gambar 2. Teknik Analisis Data

III. Hasil dan Pembahasan

Bagian ini tempat menuliskan hasil penelitian yang dijabarkan secara detail, jelas dan terurut. Hasil penelitian disajikan dalam bentuk tabel, grafik atau ilustrasi lain dan disertai dengan pembahasan yang disajikan secara terstruktur dan sistematis. Uraian performansi, kelemahan, dan kelebihan dari hasil penelitian harus dijelaskan.

Pada tahap penelitian menggambarkan beberapa proses yang terdiri dari beberapa perhitungan dan implementasi metode *naïve bayes classifier*

1) Pengumpulan Data

Pada tahap ini, ditemukan beberapa data berdasarkan kriteria dari kata-kata *body shiming* sepeti "cungkring, botak, burik, buncit, gendutan, jelek, pesek, dsb" data *tweet s* yang dikumpulkan disimpan dalam bentuk .cvs. Berikut ini daftar *tweet s* yang dikumpulkan dari Kaggle dapat dilihat pada Tabel 1 berikut:

Tabel 1. Data *tweet s Body Shaming*

ID	*Tweet* s
0	Dh panjang eh rambut kau saddiq, drpd kau botak licin sampai rambut dh panjang pon kita masih pkp lagi. Nnti kau dh ada ub…
1	jdi kek botak
4	Si botak lagi ngigau ?? https://t.co/25H4o8q8OC
106	6. diet pola makan, atur jam makan & jangan asal makan yang berlemak tinggi, awas perut buncit & sakit jantung. #alabarbar ##kulinerbandung
879	Tapi si tonggos tidak naik motor matic lhoh Mbak.. Dia naik motor pake gigi.. (bukan pake giginya sendiri maksudnya) ????????????

2) Pelabelan Data

Pelabelan data dilakukan setelah data dikumpulkan. Proses pelabelan dataset ini dilakukan dengan cara manual dengan membaca satu persatu data *tweet s* dan diberi label (positif atau negatif).

3) Implementasi Preprocessing Data

Text preprocessing adalah satu tahapan awal untuk kinerja pengolahan text yang dilakukan pada Natural Language Processing. Text preprocessing memiliki fungsi guna memudahkan dalam mengolah data, data dianalisa secara manual dengan membaca maksud dari kalimat yang ada dalam sentimen tersebut, sehingga dapat diberikan penilai[16]. Adapun langkah- langkah pada teks *preprocessing* sebagai berikut:

- *Cleaning*, digunakan untuk menghapus emotikon, *url*, *emoji*, *tag* html, tanda baca, *hashtag* dan *mention*.
- *Case Folding*, adalah mengubah semua menjadi huruf kecil, hanya huruf "a" sampai "z" yang diterima
- *Tokenizing*, adalah proses memisahkan kata-kata dari kalimat penyusunya yang nantinya disebut *token* atau *term*. *Tekenizing* yang digunakan adalah unigram yang terdiri dari satu kata. Berikut data *tweet s* yang telah melalui tahap *tekonizing* pada Tabel 2 berikut:

Tabel 2. Data *Tweet s Body Shaming*

ID	*Tweet* s	Label
0	['dh', 'panjang', 'eh', 'rambut', 'kau', 'saddiq', 'drpd', 'kau', 'botak', 'licin', 'sampai', 'rambut', 'dh', 'panjang', 'pon', 'kita', 'masih', 'pkp', 'lagi', 'nnti', 'kau', 'dh', 'ada', 'ub', '?']	Positif
1	['jdi', 'kek', 'botak']	Negatif
4	['si', 'botak', 'lagi', 'ngigau']	Positif
106	['diet', 'pola', 'makan', 'atur', 'jam', 'makan', 'amp', 'jangan', 'asal', 'makan', 'yang', 'berlemak', 'tinggi', 'awas', 'perut', 'buncit', 'amp', 'sakit', 'jantung']	Negatif
879	['tapi', 'si', 'tonggos', 'tidak', 'naik', 'motor', 'matic', 'lhoh', 'mbak', 'dia', 'naik', 'motor', 'pake', 'gigi', 'bukan', 'pake', 'giginya', 'sendiri', 'maksudnya']	Positif

- *Stopwords* adalah menghapus kata kata yang sering muncul tetapi kurang relevan dengan teks. Penghapusan kata – kata tersebut dikarenakan kata – kata tersebut tidak bermakna dan tidak memiliki pengaruh terhadap analisis sistem
- *Normalization* atau biasa disebut normalisasi adalah menormalkan kata kata yang disingkat atau yang diperpanjang menjadi kata – kata yang normal sesuai dengan Kamus Besar Bahasa Indonesia (KBBI). Berikut *tweet s* yang sudah melalui tahap normalisasi terlihat pada Tabel 3 berikut:

Tabel 3. *Tweet s* setelah *Stopword* dan *Normalization*

ID	*Tweet* s	Label
0	['rambut', 'kamu', 'saddiq', 'kamu', 'botak', 'licin', 'rambut', 'juga', 'pkp', 'nanti', 'kamu', 'ub', '?']	Positif
1	['botak']	Negatif
4	['botak', 'ngigau']	Positif
106	['diet', 'pola', 'makan', 'atur', 'jam', 'makan', 'makan', 'lemak', 'awas', 'perut', 'buncit', 'sakit', 'jantung']	Negatif

879	['tonggos', 'motor', 'matic', 'lhoh', 'mbak', 'motor', 'pakai', 'gigi', 'pakai', 'giginya', 'maksudnya']	Positif

- *Stemming* adalah mengubah kata berimbuhan dari setiap kata yang sudah disaring menjadi kata dasar.
- *Feature Model* merupakan proses penting pada klasifikasi teks untuk mengubah format tekstual yang tidak terstruktur menjadi terstruktur sehingga dapat diproses oleh algoritma *machine learning* untuk mengklasifikasikan ke kelas yang telah ditentukan. Berikut daftar *feature model* berdasarkan trigram. pada Tabel 4:

Tabel 4. Data *Tweet s* Hasil *Trigram*

ID	Frequency
nafsu makanmakan gemuk	5
kurus kerempengga nafsu	5
minum madu profesional	5
hikmah ppkm sipit	1
zoom pantat saya	1

4) Implementasi metdoe Naive Bayes Classifier

Implementasi *Naïve bayes classifier* pada data *tweet s* artinya mengimplementasikan hasil kerja dari *feature model* yang menghasilkan *feature list trigram*. Implementasi setiap *feature model* pada metode *naïve bayes classifier* memberikan tingkat performa baik akurasi, presisi, *Recall*, dan *f-measure* yang berbeda beda[17]–[19].

Proses *training classifier* menggunakan data training berupa *tweet s* yang telah diberi label kelas dan telah melalui tahap *preprocessing* yang nantinya akan menghasilkan sebuah model klasifikasi. Tahapan dalam melakukan *training data* sebagai berikut:

a) Mencari jumlah keseluruhan *tweet* s M pada data training. Misalkan data yang digunakan sebagai data training adalah data *tweet* s yang terdapat pada Tabel 9. Maka jumlah keselurah *tweet* s M = 5.

b) Mencari jumlah *tweet* s pada masing-masing label kelas. Berdasarkan data *tweet* s yang terdapat pada Tabel 9. Maka jumlah *tweet* s untuk kelas positif Npositif = 3 dan kelas negatif Nnegatif = 2.

c) Menentukan nilai prior probability tiap kelasnya P(label¬_kelas), sesuai dengan persamaan (6), maka didapatkan hasil:

P(Positif) = NPositif / M = 3/5 = 0.6

P(Negatif) = Nnegatif / M =2/5 = 0.4

d) Melakukan operasi logaritma pada nilai prior probability untuk tiap kelasnya. Hasil yang didapatkan adalah:

Log2P(Positif)=log2 0.6 = -0.73696559416621 Log2P(Negatif)=log2 0.4 = -1.3219280948874

e) Proses selanjutnya yaitu menghitung nilai jumlah kemunculan tiap fitur untuk tiap kelasnya dan jumlah total kemunculan fitur untuk tiap kelasnya. Fitur yang dimaksud adalah tiap kata yang terdapat pada *tweet* s yang telah melalui tahap preprocessing. Jumlah kemunculan fitur untuk tiap kelas dapat dilihat pada Tabel 5.

Tabel 5. Data *Tweet s* Hasil *Trigram*

Fitur	Jumlah Kemunculan	
	Positif	Negatif
Rambut	2	0
Kamu	3	0
Saddiq	1	0
Botak	2	1
Licin	1	0
Juga	1	0
Nanti	1	0

Ngigau	1	0
Diet	0	1
Pola	0	1
Makan	0	3
Atur	0	1
Jam	0	1
Lemak	0	1
Awas	0	1
Perut	0	1
Buncit	0	1
Sakit	0	1
Jantung	0	1
Tonggos	1	0
Motor	2	0
Matic	1	0
Mbak	1	0
Pakai	2	0
Gigi	1	0
Maksud	1	0
Total	**21**	**14**

f) Mencari jumlah fiitur unik B yang terdapat pada vocabulary. Jumlah fitur unik B = 26.

g) Proses selanjutnya menentukan nilai probabilitas kondisional tiap fitur untuk tiap kelas P (fitur|kelas) berdasarkan persamaan (8). Sebagai contoh nilai probabilitas kondisional untuk fitur 'bentuk' pada tiap kelasnya:

P(rambut|Positif) = (2+1)/(21+26) = 0.06382979

P(rambut|Negatif) = (0+1)/(14+26) = 0.025

h) Proses selanjutnya melakukan operasi logaritma terhadap nilai probabilitas kondisional tiap fitur untuk tiap kelas yang telah didapatkan sebelumnya.

Log2 P(rambut|Positif) = log2 0.06382979 = -3.96962628844

Log2 P(rambut|Negatif) = log2 0.025 = -5.32192809489

i) Nilai-nilai dari log prior probability untuk tiap kelas dan nilai log probabilitas kondisional tiap fitur untuk tiap kelas inilah yang nantinya dijadikan sebagai model klasifikasi. Data *Tweet* Hasil *Trigram* ditunjukkan pada Tabel 6

Tabel 6. Data *Tweet s* Hasil *Trigram*

Fitur	Jumlah Kemunculan	
	Positif	**Negatif**
Rambut	-3.96962628844	-5.32192809489
Kamu	-3.55458890217	-5.32192809489
Saddiq	-4.55458890217	-5.32192809489
Botak	-3.96962628844	-4.32192809489
Licin	-4.55458890217	-5.32192809489
Juga	-4.55458890217	-5.32192809489
Nanti	-4.55458890217	-5.32192809489
Ngigau	-4.55458890217	-5.32192809489
Diet	-5.55458856314	-4.32192809489
Pola	-5.55458856314	-4.32192809489

Makan	-5.55458856314	-3.32192809489
Atur	-5.55458856314	-4.32192809489
Jam	-5.55458856314	-4.32192809489
Lemak	-5.55458856314	-4.32192809489
Awas	-5.55458856314	-4.32192809489
Perut	-5.55458856314	-4.32192809489
Buncit	-5.55458856314	-4.32192809489
Sakit	-5.55458856314	-4.32192809489
Jantung	-5.55458856314	-4.32192809489
Tonggos	-4.55458890217	-5.32192809489
Motor	-3.96962628844	-5.32192809489
Matic	-4.55458890217	-5.32192809489
Mbak	-4.55458890217	-5.32192809489
Pakai	-3.96962628844	-5.32192809489
Gigi	-4.55458890217	-5.32192809489
Maksud	-4.55458890217	-5.32192809489
Log P(kelas)	**-1.2607945727217**	**-1.2537013046714**

Setelah pembentukan model klasifikasi selesai dilakukan, maka proses klasifikasi siap untuk dilakukan. Misalkan sebuah *tweet* s yang berisi kalimat 'Kamu Botak Licin' yang akan ditentukan sentimennya. Maka tahap untuk melakukan proses klasifikasi pada kalimat tersebut adalah:

- *Tweet* s melalui tahap preprocessing, hasil preprocessing berdasarkan kalimat 'Kamu Botak Licin' adalah 'kamu botak licin'.
- Menentukan nilai log probabilitas kondisional setiap fitur yang terdapat pada *tweet* s hasil preprocessing 'kamu botak licin' dengan mengambil nilai log probabilitas kondisional setiap fitur untuk setiap kelas yang terdapat pada model klasifikasi. Hasilnya dapat dilihat pada Tabel. Mengambil nilai log prior probability tiap kelas log P(kelas) pada model klasifikasi.
- Penentuan kelas sentiment berdasarkan persamaan (5) dengan mencari nilai maksimum dari nilai probabilitas akhir tiap kelas P(kelas|*tweet* s)

P(Positif | kamu botak licin) = -1.2067945727217 + -3.55458890217 + -3.96962628844 + -4.55458890217

= -13.2855986655017

P(Negatif | kamu botak licin) = -1.2537013046714 + -5.32192809489 + -4.32192809489 + -5.32192809489

= -16.2194855893414

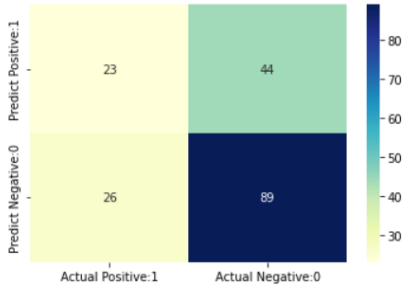
Cmap = max(P(Positif|'kamu botak licin'), P(Negatif|'kamu botak licin'))

= max (-13.2855986655017, -16.2194855893414)

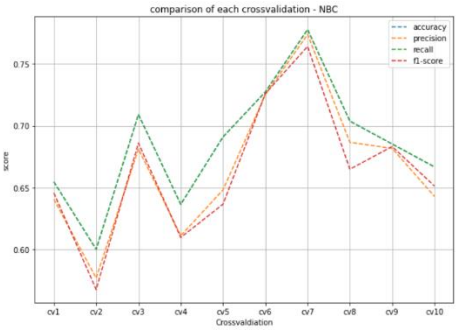
= -13.2855986655017

Tweet s 'Kamu Botak Licin' diklasifikasikan dalam analisis sentiment kelas positif.

j) Berdasarkan pengujian performa metode *Naïve bayes classifier* pada data *tweet* s body shaming dengan menggunakan pengujian performa akurasi, presisi, *Recall* dan *f-measure*, maka didapatkan hasil pengujian yang akan diuraikan secara jelas sebagai berikut. Crossvalidation dilakukan pada tahap implemetasi metode klasifikasi agar memperoleh berbagai macam nilai yang natinya dapat divisualisasikan dalam boxplot[20], [21], [30]–[39], [22]–[29].



Gambar 3. *Consudion Matrix*



Gambar 4. Grafik Pengujian Cross Validation Model Trigram

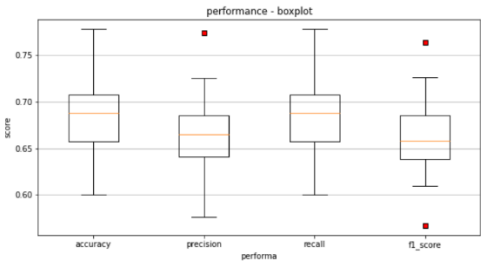
Berikut nilai performa akurasi, presisi, *Recall*, dan *f-measure* dengan *feature* model trigram menggunakan metode *naïve bayes classifier* ditujukan pada Tabel 5. Data *tweets* hasil trigram ditunjukkan pada Tabel 7 dan Nilai pengujian *cross validation* ditunjukkan pada Tabel 8. Nilai *boxplot model trigram* ditunjukkan pada Gambar 5.

Tabel 7. Data *Tweet s* Hasil *Trigram*

Feature model	**Performa**			
	Accuracy	***Precision***	***Recall***	***F-measure***
Tiagram	0.68515151	0.66673666	0.68515151	0.66342268

Tabel 8. Nilai Pengujian *Cross Validation*

Feature model	**Iterasi**	**Performa**			
		Accuracy	***Precision***	***Recall***	***F-measure***
Tiagram	1	0.65693431	0.63042922	0.65693431	0.60910641
	2	0.65693431	0.57668127	0.65693431	0.58453306
	3	0.72794118	0.72107803	0.72794118	0.71717377
	4	0.70588235	0.68381602	0.70588235	0.68391713



Gambar 5. Nilai *Boxplot Model Trigram*

5) Pembahasan

Dataset yang digunakan pada penelitian ini adalah data yang didapat dari situs kaggle. Pada dataset ini terdiri dari 729 sampel dan setelah melalui proses *preprocessing* data, maka data menjadi 728 dengan menghapus data *tweet s* yang sama.

Langkah-langkah implementasi pada aplikasi terdiri dari beberapa *library*/fungsi/fitur yaitu deklarasi *library* analisis data yang memiliki struktur data dalam bentuk tabel, deklarasi *library* untuk mengolah data berupa teks, deklarasi *library* untuk memuat kata-kata yang mengandung *stopword* atau kata kurang penting dalam sebuah kalimat, deklarasi *library* untuk memisahkan kata per kata pada sebuah kalimat, deklarasi *library* menghitung berapa banyak kata yang muncul pada sebuah dataset, deklarasi *library* untuk melakukan proses penggunaan kata dasar, deklarasi *library* untuk fitur ektrasi, deklarasi *library* untuk *regular expression*, deklarasi untuk mencari string atau teks dengan pola tertentu, deklarasi *library* untuk mengelola data berupa string, deklarasi *library* untuk mengurangi kata-kata yang terinfleksi dalam bahasa Indonesia, deklarasi *library* untuk membagi dataset menjadi data training dan data *testing*, deklarasi *library* untuk menerapkan nilai K-Fold dalam sebuah metode, deklarasi *library* untuk menerepkan cross validation dalam sebuah metode, deklarasi *library* untuk metode naïve bayes classifier, deklarasi *library* untuk mencari nilai *Accuracy*, deklarasi *library* untuk memvisualisasikan nilai *Precision*, *Recall* dan *f-measure*, deklarasi *library* untuk menganalisis nilai actual dan nilai prediksi, deklarasi *library* untuk menvisualisasikan sebuah data, dan deklarasi *library* untuk memvisualisasikan sebuah plot atau grafik. Juga terdapat komponen seperti dataset, Sistem, Sistem operasi.

Sedangkan untuk proses pambacaan dataset dilakukan dengan beberapa tahap yaitu membaca file CSV berupa data *tweet* s, membuat sebuah fungsi *python* dengan parameter text, menampilkan semua kata yang mengandung karakter '@' didepannya, menampilkan semua kata yang mengandung karakter '#' didepannya, mengganti semua kata yang mengandung karakter '@' didepannya agar bisa dihapus, mengganti semua kata yang mengandung karakter '#' didepannya agar bisa dihapus, menghapus semua kata yang didepannya terdapat karakter '@' atau '#', dan menjalankan fungsi untuk Menghapus semua kata yang didepannya terdapat karakter '@' atau '#'.

Kfold pada cross validation yang digunakan yakni 10, menghitung performa *Accuracy* menggunakan metode *naïve bayes classifier* dengan data training dan data testing berdasarkan data training, menghitung performa *Precision* menggunakan metode naivve bayes classifier dengan data training dan data testing berdasarkan data training, menghitung performa *Recall* menggunakan metode *naïve bayes classifier* dengan data training dan data testing berdasarkan data training, menghitung performa *f-measure* menggunakan metode *naïve bayes classifier* dengan data training dan data testing berdasarkan data training, menampilkan nilai rata-rata *Accuracy*, menampilkan nilai rata-rata *Precision*, menampilkan nilai rata-rata *Recall*, dan menampilkan nilai rata-rata *f-measure*

IV. Kesimpulan

Berdasarkan hasil analisis yang telah dilakukan, maka dapat disimpulkan bahwa performa akurasi, presisi, *recall*, dan f-*measure* dengan feature model trigram menggunakan metode *naïve bayes classifier* pada dataset *tweets body sh*aming maka didapatkan hasil akurasi 61%, presisi 56%, *recall* 55%, dan f-*measure* 55%. Berdasarkan hasil tersebut dapat dikatakan bahwa performa dari penggunaan metode *naïve bayes* terhadap dataset tersebut tidak cukup baik karena sulitnya didapatkan data yang seimbang sehingga model yang dihasilkan pada bagian presisi terdapat nilai performa yang *outlier.*

Daftar Pustaka

[1] F. Nurhuda, S. Widya Sihwi, and A. Doewes, "Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier," *J. Teknol. Inf. ITSmart*, vol. 2, no. 2, p. 35, 2016, doi: 10.20961/its.v2i2.630.

[2] S. R. I. Rezeki, "Penggunaan Sosial Media Twitter dalam Komunikasi Organisasi (Studi Kasus Pemerintah Provinsi Dki Jakarta Dalam Penanganan Covid-19)," *J. Islam. Law Stud.*, vol. 04, no. 02, pp. 63–78, 2020.

[3] A. Tamaraya and D. Ubaedullah, "Dampak Penggunaan Twitter Terhadap Pengungkapan Diri Mahasiswa," *Interak. Perad. J. Komun. dan Penyiaran Islam*, vol. 1, no. 1, pp. 29–37, 2021, doi: 10.15408/interaksi.v1i1.20878.

[4] S. Sari, U. Khaira, P. Pradita, and T. S. Tri, "… Beauty Shaming Di Media Sosial Twitter Menggunakan Algoritma SentiStrength: Sentiment Analysis Against Beauty Shaming Comments on Twitter Social Media …," *Indones. J. …*, vol. 1, no. 1, pp. 71–78, 2021.

[5] T. F. Fauzia and L. R. Rahmiaji, "Memahami pengalaman Body Shaming pada remaja perempuan," pp. 4–5, 2019.

[6] K. Fitria and Y. Febrianti, "the Interpretation and Attitude of Body Shaming Behavior on Social Media (a Digital Ethnography Study on Instagram)," *Diakom J. Media dan Komun.*, vol. 3, no. 1, pp. 12–25, 2020, doi: 10.17933/diakom.v3i1.78.

[7] W. N. Putri, A. E. Budiwaspada, and D. Ratri, "REKOMENDASI RANCANGAN KAMPANYE SOSIAL TEMA BODY SHAMING BAGI GENERASI Z," *J. Komun. Vis. Wimba*, vol. 12, no. 2, pp. 110–123, 2021.

[8] D. Geofani, "Pengaruh cyberbullying body shaming pada media sosial instagram terhadap kepercayaan diri wanita karir di Pekanbaru," *Jom Fisip*, vol. 6, pp. 2–6, 2019.

[9] C. Pricilia, D. Yoanita, and D. Budiana, "Pengaruh Bodily Shame di Instagram terhadap Konsep Diri Remaja Perempuan," *J. E-Komunikasi*, vol. 7, no. 2, pp. 1–12, 2019.

[10] M. Hall, "A decision tree-based attribute weighting filter for Naive Bayes," *Res. Dev. Intell. Syst. XXIII - Proc. AI 2006, 26th SGAI Int. Conf. Innov. Tech. Appl. Artif. Intell.*, pp. 59–70, 2007, doi: 10.1007/978-1-84628-663-6_5.

[11] Bustami, "Penerapan Algoritma Naive Bayes," *J. Inform.*, vol. 8, no. 1, pp. 884–898, 2014.

[12] R. P. Pratama, I. Werdiningsih, and I. Puspitasari, "Sistem Pendukung Keputusan Pemilihan Siswa Berprestasi di Sekolah Menegah Pertama dengan Metode VIKOR dan TOPSIS," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 3, no. 2, p. 122, 2017, doi: 10.20473/jisebi.3.2.122-128.

[13] D. N. Sari, D. N. Sari, F. Adelia, F. Rosdiana, B. B. Butar, and M. Hariyanto, "Analisa Sentimen Terhadap Review Produk Kecantikan Menggunakan Metode Naive Bayes Classifier," *JIKA (Jurnal Inform.*, vol. 4, no. 3, p. 109, 2020, doi: 10.31000/jika.v4i3.3086.

[14] F. F. Mailo and L. Lazuardi, "Analisis Sentimen Data Twitter Menggunakan Metode Text Mining Tentang Masalah Obesitas di Indonesia," *J. Inf. Syst. Public Heal.*, vol. 4, no. 1, pp. 28–36, 2019.

[15] J. H. Jaman, Hannie, and M. S. Simatupang, "Sentiment Analysis of the Body-Shaming Beauty Vlog Comments," *Proc. 7th Math. Sci. Comput. Sci. Educ. Int. Semin. MSCEIS 2019*, 2020, doi: 10.4108/eai.12-10-2019.2296530.

[16] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New Avenues in Opinion Mining and Sentiment Analysis," *IEEE Comput. Soc.*, no. April, pp. 15–21, 2013.

[17] H. Azis, F. T. Admojo, and E. Susanti, "Analisis Perbandingan Performa Metode Klasifikasi pada Dataset Multiclass Citra Busur Panah," *Techno.Com*, vol. 19, no. 3, 2020.

[18] Herman *et al.*, "Comparison of Artificial Neural Network and Gaussian Naïve Bayes in Recognition of Hand-Writing Number," *Proc. - 2nd East Indones. Conf. Comput. Inf. Technol. Internet Things Ind. EIConCIT 2018*, no. 1, pp. 276–279, 2018, doi: 10.1109/EIConCIT.2018.8878651.

[19] A. Fitria and H. Azis, "Analisis Kinerja Sistem Klasifikasi Skripsi menggunakan Metode Naïve Bayes Classifier," *Pros. Semin. Nas. Ilmu Komput. dan Teknol. Inf*, vol. 3, pp. 102–106, 2018.

[20] F. Tangguh and Y. Islami, "Analisis performa algoritma Stochastic Gradient Descent (SGD) dalam mengklasifikasi tahu berformalin," *Indones. J. Data Sci.*, vol. 3, no. 1, pp. 1–8, 2022.

[21] A. I. Auliyah, "Implementasi Kombinasi Algoritma Enkripsi Rivest Shamir Adleman (Rsa) dan Algoritma Kompresi Huffman Pada File Document," *Indones. J. Data Sci.*, vol. 1, no. 1, pp. 23–28, 2020.

[22] L. B. C. Tanujayaa, B. Susanto, and A. Saragiha, "Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Fitur Mode Audio Spotify," *Indones. J. data Sci.*, vol. 1, no. 3, pp. 68–78, 2020.

[23] Wahyu ngestisari, "The Perbandingan Metode ARIMA dan Jaringan Syaraf Tiruan untuk Peramalan Harga Beras," *Indones. J. Data Sci.*, vol. 1, no. 3, pp. 96–107, 2020, doi: 10.33096/ijodas.v1i3.18.

[24] A. A. D. Halim and S. Anraeni, "Analisis Klasifikasi Dataset Citra Penyakit Pneumonia menggunakan Metode K-Nearest Neighbor (KNN)," *Indones. J. Data Sci.*, vol. 2, no. 1, pp. 01–12, 2021, doi: 10.33096/ijodas.v2i1.23.

[25] D. Cahyanti, A. Rahmayani, and S. Ainy, "Analisis performa metode Knn pada Dataset pasien

pengidap Kanker Payudara," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 39–43, 2020.

[26] Sugiarti and Mirnawati, "Implementasi Algoritma Goverment Standard (GOST) dalam Pengamanan File Dokumen," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 52–56, 2020.

[27] H. Nursan and Muslim, "Penerapan Metode Digital Watermarking dan Privilege pada Dokumen Skripsi," *Indones. J. Data Sci.*, vol. 1, no. 1, pp. 19–22, 2020.

[28] F. T. Admojo and Ahsanawati, "Klasifikasi Aroma Alkohol Menggunakan Metode KNN," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 34–38, 2020.

[29] N. Litha and T. Hasanuddin, "Analisis Performa Metode Moving Average Model untuk Prediksi Jumlah Penderita Covid-19," *Indones. J. Data Sci.*, vol. 1, no. 3, pp. 87–95, 2020.

[30] Harlinda and Nasir, "Perancangan sistem pendukung keputusan dalam pengalokasian dana bantuan sosial di kabupaten pinrang dengan menggunakan metode AHP," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 44–51, 2020.

[31] S. Sahar, "Analisis Perbandingan Metode K-Nearest Neighbor dan Naïve Bayes Clasiffier Pada Dataset Penyakit Jantung," *Indones. J. Data Sci.*, vol. 1, no. 3, pp. 79–86, 2020, doi: 10.33096/ijodas.v1i3.20.

[32] Muh Syawal, P. L. L. Belluano, and A. R. Manga, "Implementasi Algoritma Genetika Untuk Penjadwalan Laboratotium Fakultas Ilmu Komputer Universitas Muslim Indonesia," *Indones. J. Data Sci.*, vol. 2, no. 1, pp. 29–37, 2021, doi: 10.33096/ijodas.v2i1.29.

[33] H. Azis, "Klasifikasi Penyakit Jantung Menggunakan Metode K-Nearest Neighbor," *Indones. J. Data Sci.*, pp. 1–4, 2018, doi: https://doi.org/10.33096/ijodas.v1i1.3.

[34] A. Maulida, "Penerapan Metode Klasifikasi K-Nearest Neigbor pada Dataset Penderita Penyakit Diabetes," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 29–33, 2020.

[35] Hasran, "Klasifikasi Penyakit Jantung Menggunakan Metode K-Nearest Neighbor," *Indones. J. Data Sci.*, vol. 1, no. 1, pp. 1–4, 2020.

[36] I. P. Putri, "Analisis Performa Metode K- Nearest Neighbor (KNN) dan Crossvalidation pada Data Penyakit Cardiovascular," *Indones. J. Data Sci.*, vol. 2, no. 1, pp. 21–28, 2021, doi: 10.33096/ijodas.v2i1.25.

[37] D. Susanti, "Analisis Modifikasi Metode Playfiar Cipher Dalam Pengamanan Data," *Indones. J. Data Sci.*, vol. 1, no. 1, pp. 1–80, 2020.

[38] A. Prasetya Wibawa, W. Lestar, A. Bella Putra Utama, I. Tri Saputra, and Z. Nabila Izdihar, "Multilayer Perceptron untuk Prediksi Sessions pada Sebuah Website Journal Elektronik," *Indones. J. Data Sci.*, vol. 1, no. 3, pp. 57–67, 2020, doi: 10.33096/ijodas.v1i3.15.

[39] M. I. Maulana, "Implementasi Bot Telegram pada Proses Retrieval Data dalam Database," *Indones. J. Data Sci.*, vol. 2, no. 1, pp. 13–20, 2021, doi: 10.33096/ijodas.v2i1.24.