

Analisis Sentimen Komentar Netizen Pada Brand Skincare The Originote Menggunakan Metode Naïve Bayes

¹Nuranna Muslimah, ^{2*}Agung Sutikno, S.Kom, M.M.S.I
Sistem Informasi, Teknik dan Ilmu Komputer, Universitas Nahdlatul Ulama Indonesia
Jakarta 10320, Indonesia
[1nurannamuslimah@gmail.com](mailto:nurannamuslimah@gmail.com), [2*sagung@unusia.ac.id](mailto:sagung@unusia.ac.id)

Received on 30-10-2023, revised on 11-11-2023, accepted on 15-12-2023

Abstrak

Media sosial pada masa kini merupakan media komunikasi yang paling banyak digunakan dan sangat populer pada masa sekarang. Salah satu media sosial yang sedang banyak diminati saat ini adalah *instagram*. Dengan media sosial *Instagram*, informasi warganet sangat beragam, informasi yang berada pada *Instagram* tersebut mempunyai data yang dapat diolah menjadi Analisa sentiment. Tujuan penelitian ini adalah untuk mengetahui megetahui review warganet terhadap brand skincare *theoriginote* dengan menggunakan algoritma *naïve bayes classifier* Dalam menganalisis sentiment dengan menggunakan aplikasi *rapidminer*. Hasil penelitian ini mengungkapkan keluaran sistem berupa visualisasi data sentiment positif dan negative pada *tools rapidminer*. Pada penelitian ini mendapatkan performance berupa hasil *akurasi* 78,48% *class precission* 71,98% *class recal* 93,30% dan nilai *F1 Score* 80,00% hasil tersebut dapat disimpulkan bahwa *algoritma naïve bayes* dapat menganalisis sentiment dengan baik oleh sistem. Pada penelitian ini dapat memberikan masukan pada brand local skincare *theoriginote* dalam merilis produk baru yang mereka punya.

Kata Kunci: Analisis Sentimen, Skincare, Algoritma Naïve Bayes, Rapidminer

I. PENDAHULUAN

Media sosial merupakan alat komunikasi diantaranya seperti Twitter, Facebook, Instagram dan juga Youtube yang sangat banyak diminati oleh masyarakat, seiring berjalannya waktu masyarakat banyak menggunakan media sosial sebagai alat promosi diantaranya seperti promosi produk, promosi politik, dan juga mempromosi hasil karya tulisan semua itu dilakukan untuk meningkatkan popularitas. Salah satunya media sosial yang telah dimanfaatkan dalam mempromosikan adalah Instagram.

Instagram adalah suatu *platform media sosial* yang memproitaskan berbagai gambar dan rekaman video. *Platform* ini pertama kali diluncurkan pada bulan Oktober 2010 *isntagram* diciptakan oleh Kevin Systrom dan Mike Krieger, dan sejak saat itu telah menjadi salah satu platform sosial media yang paling terkenal di

seluruh dunia. Instagram memungkinkan penggunanya untuk membuat profil pribadi, mengunggah foto dan video, serta berinteraksi dengan pengguna lain melalui fitur-fitur seperti komentar, suka (like), dan pesan langsung. Adapun objek pada penelitian ini adalah dengan menggunakan komentar atau opini Masyarakat yang berada pada postingan *Instagram* pada brand *The Originote*. Opini merupakan pendapat, gagasan atau pemikiran yang dimaksudkan untuk menggambarkan kecenderungan atau pilihan khusus terhadap pandangan dan ideologi, namun tidak obyektif karena belum mendapat konfirmasi atau pengalaman pengujian, dapat pula berupa pernyataan tentang sesuatu yang akan diterapkan dalam suatu hal masa depan dan kebenarannya atau salah dan tidak dapat ditentukan secara langsung, misalnya dengan pembuktian dengan induksi [1]. Salah satu media sosial yang paling populer di kalangan masyarakat Indonesia saat ini adalah Instagram. Di aplikasi jejaring sosial Instagram, Anda bisa melakukan banyak hal, termasuk mempromosikan produk atau jasa yang kita tawarkan.

Dalam mempromosikan suatu hal pada media sosial salah satunya adalah instagram dilihat dari opini masyarakat pada komentar yang mereka ajukan hal tersebut maka dapat dilihat apa karya atau produk yang dipromosikan layak atau tidak. Adapun objek pada penelitian ini adalah instagram pada brand kecantikan skincare yaitu *The Originote* karena banyaknya produk yang telah dirilis oleh brand tersebut maka penelitian ini berfungsi untuk menilai *review* masyarakat pada brand tersebut. Tentu saja informasi yang terkandung pada *instagram* akan sangat berarti sebagai sarana untuk penentu kebijakan pada penelitian ini dan bisa dilakukan dengan text mining.

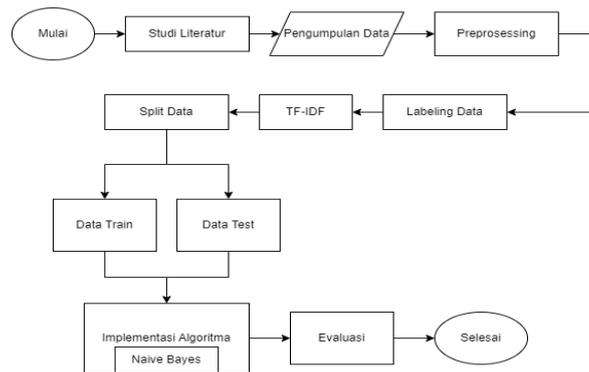
Text mining digunakan untuk menggali informasi yang bermanfaat dari serangkaian dokumen yang datanya tidak terstruktur dalam bentuk teks. Pada *Text Mining*, Langkah pengambilan informasi dapat menghasilkan analisis sentiment yang membantu dalam menentukan secara *erosional* apakah suatu pernyataan bersifat positif atau negatif. Tujuan Text Mining adalah untuk mengolah dokumen yang bersifat tidak berstruktur atau semi terstruktur. Proses Text Mining secara efisien mengekstrak informasi yang relevan dari beberapa dokumen diagram [2]. Tujuan digunakannya *text mining* dalam penelitian ini dengan menggunakan analisis sentiment. Analisis sentimen adalah proses mengidentifikasi sentimen dan mengelompokkan kutub-kutub teks dalam suatu dokumen atau kalimat sehingga dapat diidentifikasi kategori sentimen positif dan negatif. Saat ini, para peneliti banyak menggunakan analisis sentimen sebagai cabang penelitian ilmu komputer. Jejaring sosial, seperti Instagram dan Twitter, sering digunakan dalam analisis sentimen untuk menentukan opini publik. Analisis sentimen juga dapat dibandingkan dengan analisis opini karena berfokus pada opini positif atau negatif

Penelitian ini mengusulkan penerapan algoritma *naïve bayes* dalam menentukan sentimen pada komentar *instagram* dan menggunakan aplikasi *rapidminer* sebagai *tools* untuk penunjang dalam memvalidasi data yang telah diolah. Hal ini dikarenakan algoritma *naïve bayes* dapat memproses data dengan jumlah yang sangat banyak.

II. METODOLOGI PENELITIAN

A. Tahap Penelitian

Pada tahap penelitian ini menggunakan analisis sentiment dengan metode algoritma *naïve bayes*. Data yang telah diolah akan dilakukan validasi aturan pada aplikasi Rapid Miner 10.1. Adapun tahap-tahap pada penelitian yang dilakukan yaitu sebagai berikut :



Gambar 1. Tahap Penelitian

Penjelasan dari gambar 1.1 sebagai berikut :

1. Studi Literatur, pada tahap ini penulis mengumpulkan jurnal-jurnal sebagai penelitian
2. Pengumpulan Data. Data yang diperoleh pada penelitian ini dilakukan dengan cara *export* pada web <https://exportcomments.com>
3. *Preprocessing*. Pada tahap ini data yang telah dikumpulkan lalu dilakukan pembersihan dengan menggunakan tahapan yang ada pada *preprocessing*.
4. Pelabelan Data, data yang telah dibersihkan kemudian dilabeli secara manual
5. Implementasi algoritma, pada tahap ini dataset yang sudah dilabeli kemudian dibagi menjadi dua pembagian data yaitu data latih dan data uji setelah itu dilakukan implementasi *naïve bayes*.
6. Evaluasi, hasil dari implementasi algoritma *naïve bayes* berupa akurasi.

B. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data yang diperoleh pada postingan komentar Instagram pada akun brand The Originote dan di *export* dengan menggunakan web <https://exportcomments.com> lalu di kumpulkan pada file excel.

C. *Preprocessing*

Preprocessing merupakan langkah terpenting untuk langkah selanjutnya yaitu reduksi atribut yang mempengaruhi proses klasifikasi. Data yang diperoleh pada tahap ini dimasukkan sebagai data mentah dan data kotor, sehingga diperoleh hasil dari proses tersebut data berkualitas yang akan memudahkan proses klasifikasi. *Preprocessing* ini mencakup beberapa langkah yaitu *cleansing*, *transform cases*, *tokenize*, *stopword*, dan filter token.

D. Algoritma Naïve Bayes

Algoritma Naïve Bayes adalah salah satu metode yang sangat umum digunakan dalam *data mining* karena kepraktisannya, kecepatan pemrosesan yang tinggi, implementasi yang mudah berkat strukturnya yang sederhana, serta tingkat efektivitas yang tinggi.[3] *Naïve Bayes* adalah suatu *teorema* yang berhubungan dengan konsep probabilitas bersyarat. Secara umum, *Teorema Bayes* dapat dirumuskan pada persamaan seperti yang berikut ini [4]:

$$P(A \setminus B) = \frac{P(A \setminus B)P(A)}{P(B)}$$

$P(V_j)$ dan probabilitas kata a_i untuk setiap kategori $P(a_i|V_j)$ dihitung pada saat training yang dirumuskan sebagai berikut :

$$P(V_j) = \frac{\text{Dokumen}_j}{\text{training}}$$

$$P(a_i|V_j) = \frac{n_{i+1}}{n + \text{kosakata}}$$

Keterangan :

Dokumen_j = Jumlah dokumen pada kategori j

training = Jumlah dokumen yang digunakan dalam proses training

n_i = jumlah kemunculan kata a_i pada kategori V_j

n = Jumlah kosakata yang muncul pada kategori V_j

kosakata = Jumlah kata unik pada semua data training

E. Evaluasi

Evaluasi performance merupakan hasil dari klasifikasi untuk mengukur nilai performance pada sistem yang telah dibuat. Hasil dari evaluasi untuk pengujian yaitu *accuracy*, nilai *precession*, dan nilai *recal*. Pada sistem akurasi bagus untuk digunakan jika positif dan negative seimbang, sedangkan *F1 score* baik digunakan pada distribusi positif dan negative tidak seimbang seperti *false positif* dan *false negative* yang angkanya masih tinggi.

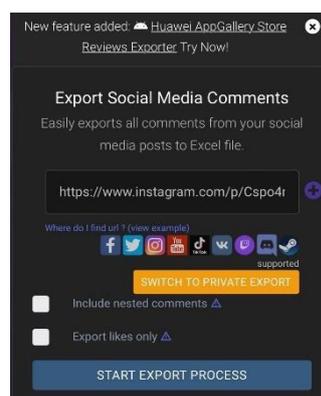
$$\text{Accuracy} = \frac{TP + TN}{(TP + TN + FP + FN)}$$

$$\text{F 1 Score} = 2 \times \frac{\text{Precession} * \text{Recall}}{\text{Precession} + \text{Recall}}$$

III. HASIL DAN PEMBAHASAN

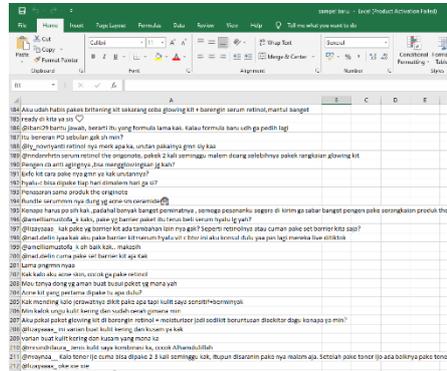
F. Pengumpulan Data

Data pada penelitian ini merupakan data komentar pada postingan Instagram *Brand The Originote* pada akun *@theoriginote* dengan postingan empat bulan belakang dengan menghasilkan 1.279 komentar pada postingan tersebut dengan menggunakan website <https://exportcomments.com> dengan memasukkan link pada postingan the originate.



Gambar 2. Halaman Export Komen

Data yang sudah diperoleh pada *website* <https://exportcomments.com> dan disimpan pada *format file excel csv* akan dilakukan pembersihan data yang dimana data atau kata yang tidak diperlukan akan dihilangkan dalam *preprocessing*. Hasil dari data yang sudah diperoleh dengan *format excel.csv*

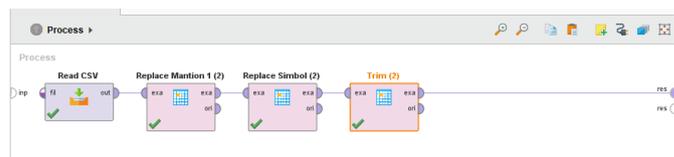


Gambar 3. Format File CSV

G. RapidMiner

RapidMiner adalah *software tools open source* yang berfungsi untuk *data mining*. *RapidMiner* diperasikan pada sebuah lingkungan untuk *machine learning, data mining, text mining, predictive analytics* [5]. Perangkat ini dapat digunakan dalam konteks aplikasi bisnis dan komersial, juga bermanfaat dalam bidang penelitian, Pendidikan, pelatihan, *prototyping* cepat, serta pengembangan aplikasi. *Rapidminer* mendukung seluruh tahapan dalam proses pembelajaran mesin, termasuk persiapan data, *visualisasi* hasil, *validasi*, dan pengoptimalan. *Rapidminer* dikembangkan menggunakan model *open source* [6].

H. Cleansing



Gambar 1. Operator Cleansing

Pada tahap *cleansing*, data yang sudah diperoleh yang berupa dokumen untuk dibersihkan kata yang tidak dapat digunakan sama terhadap klasifikasi analisis sentimen. Dokument pada komentar Instagram banyak memiliki atribut yang tidak berpengaruh terhadap sentiment. Contoh dari atribut yang tidak memiliki pengaruh yaitu mention ('@ ') dan karakter symbol (~!@#%\$%^&* ?,:.{}|[]). Atribut yang tidak berpengaruh tersebut akan dibersihkan dari dokument.

Tabel I. Ilustrasi Cleansing

Input Proses	Output Proses
--------------	---------------

@theoriginote Belum pake seminggu, muka udah keliatan perbedaannya,bawaanya jadi pen ngaca terus..luv banget monangiss semoga cocok terus ya tapi kmren po nya beneran sebulan	Belum pake seminggu, muka udah keliatan perbedaannya,bawaanya jadi pen ngaca terus..luv banget monangiss semoga cocok terus ya tapi kmren po nya beneran sebulan
--	--

I. Transform Cases



Gambar 2. Operator Preprocessing

Disini Transform Cases adalah langkah untuk mengubah data dari huruf besar ke huruf kecil. Tujuannya adalah untuk menghapus data duplikat yang hanya berbeda hurufnya saja.

Tabel II. Ilustrasi Transform Cases

Output Proses	Input Proses
Belum pake seminggu, muka udah keliatan perbedaannya,bawaanya jadi pen ngaca terus..luv banget monangiss semoga cocok terus ya tapi kmren po nya beneran sebulan	belum pake seminggu, muka udah keliatan perbedaannya,bawaanya jadi pen ngaca terus..luv banget monangiss semoga cocok terus ya tapi kmren po nya beneran sebulan

J. Tokenize

Langkah *Tokenisasi* melibatkan pemisahan kata berdasarkan setiap kata dan mengurutkannya menjadi beberapa bagian. Kata-kata dalam dokumen yang dimaksud adalah kata-kata yang dipisahkan spasi sehingga proses perolehannya melibatkan memasukkan kata-kata tersebut ke dalam database dan kemudian memberikan bobot padanya.

Tabel III. Ilustrasi Tokenize

Output Proses	Input Proses
banyakin konten seperti ini kak, kalo bisa semua produknya deh wkwk biar kita ga bingung lagi dan aman make semua produknya.	'banyakin' 'konten' 'seperti' 'ini' 'kalo' 'bisa' 'semua' 'produknya' 'deh' 'wkwk' 'biar' 'kita' 'ga' 'bingung' 'lagi' 'dan' 'aman' 'make' 'semua' 'produknya'

K. Stopword

Langkah stopword merupakan langkah menghilangkan kata-kata yang tidak sesuai dengan topik dokumen, jika kata-kata tersebut tidak mempengaruhi keakuratan klasifikasi sentimen dokumen. Kata tersebut perlu dihapus dari kumpulan kata sandi di database.

Tabel IV. Ilustrasi Stopword

Output Proses	Input Proses
belum pake seminggu, muka udah keliatan perbedaannya,bawaanya jadi pen ngaca terus..luv banget monangiss semoga cocok terus ya tapi kmren po nya beneran sebulan	pake seminggu, muka udah keliatan perbedaannya,bawaanya jadi ngaca terus banget monangiss semoga cocok kmren po nya beneran sebulan

L. Filter Token

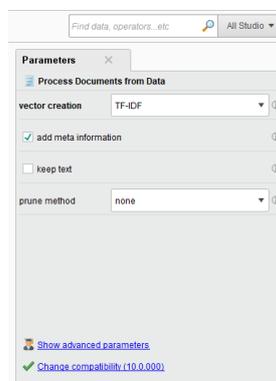
Pada tahap ini merupakan tahap untuk menghilangkan kata yang terdapat pada dokumen yang merupakan kata singkatan yang tidak berpengaruh pada klasifikasi sentiment.

Tabel V. Ilustrasi Filter Token

Output Proses	Input Proses
Yg order produk ini harus super duper sabar	order produk ini harus super duper sabar

M. TF-IDF

IDF adalah nilai kebalikan dari nilai DF. Dengan mengalikan TF dengan IDF, kita mendapatkan bobot kata tersebut. Bobot kata akan lebih tinggi jika kata tersebut sering muncul dalam suatu dokumen dan lebih rendah jika muncul di banyak dokumen [7]



Gambar 3. Parameter TF-IDF

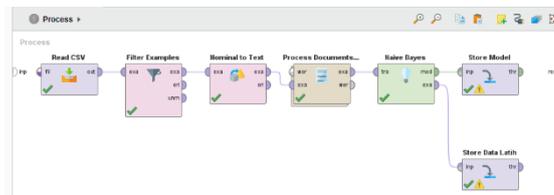
Tahap pembobotan TF-IDF yaitu dengan menentukan kata yang paling banyak muncul dengan lebih sederhana dimana data yang telah diolah akan di visualisasikan nilai kata yang paling banyak dipakai dan di ambil 25 kata pemakaian terbanyak.



Gambar 7. Visualisasi Wordcloud

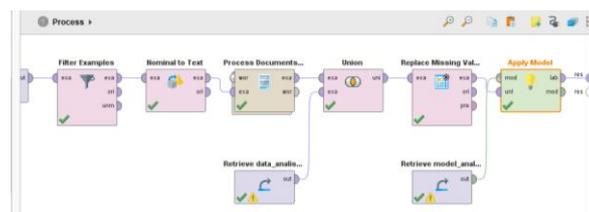
N. Implementasi Algoritma

Pada tahap pertama dalam *implentasi algoritma Naïve Bayes* adalah dengan membagi dua data yaitu data latih dan data testing yang telah kita olah yang sudah melalui *preprosessig* dan data training sudah dilabeli secara manual akan di pisahkan dengan data testing yang belum memiliki label. Selanjutnya, masukkan *operator-operator* yang digunakan pada proses tahap membagi dua data yaitu data training dan data testing.



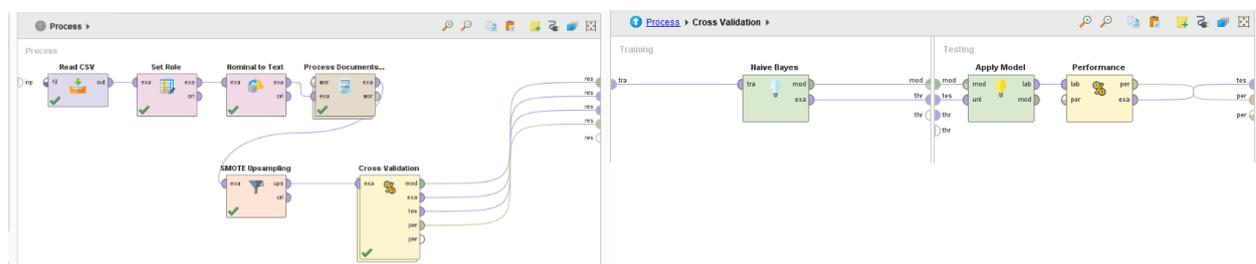
Gambar 4. Operator Split Data

Selanjutnya merupakan tahap data testing yaitu dengan melabeling menggunakan *tools rapidminer* yang sebelumnya pemisahan dilakukan menjadi dua bagian, yaitu data latih dan data uji, selanjutnya diterapkan menggunakan *Algoritma Naïve Bayes*. yang nantinya algoritma tersebut akan belajar bagaimana label sentiment positif dan negative melalui data training.



Gambar 5. Operator Labeling Data

Setelah mendapatkan hasil data testing yang sudah dilabeli sentiment positif dan negative dengan menggunakan algoritma *Naïve Bayes* pada *tools Rapidminer*. Selanjutnya adalah menentukan *performance* dari hasil data testing yang sudah dilabeli secara otomatis pada aplikasi rapidminer dengan menggunakan algoritma *Naïve Bayes*.



Gambar 6. Operator Performance

Hasil pengujian menggunakan *Naïve Bayes* dengan aplikasi *Rapidminer* akan mendapatkan *accuracy*, *class precesion* dan *class recal*. Berikut merupakan hasil dari perhitungan *performance*.

accuracy: 78.48% +/- 7.87% (micro average: 78.49%)			
	true positif	true negatif	class precision
pred. positif	167	65	71.98%
pred. negatif	12	114	90.48%
class recall	93.30%	83.69%	

Gambar 7. Hasil Performance

Tabel VI. Confusion Matrix

		True Class	
		Positive	Negative
Predicted Class	Positive	167	65
	Negative	12	114

$$\text{Accuracy} = \frac{167 + 114}{(167 + 114 + 65 + 12)} \times 100\% = 78.49\%$$

$$F1 \text{ Score} = 2 \times \frac{0,71 * 0,93}{0,71 + 0,93} \times 100 = 80.00\%$$

Dapat dilihat pada pengujian dengan menggunakan *Rapidminer* hasil yang diperoleh adalah *accuracy* 78,49%, *class precision* 71,98% dan *class recal* 93,30% hasil tersebut merupakan hasil dari pengujian dengan menggunakan metode *Naïve Bayes* pada aplikasi *Rapidminer*. Pada hasil dari *performance* masih memiliki angka yang besar pada *False Positif* dan *False Negatif*, maka akan dihitung menggunakan *F1 Score* dengan menghasilkan nilai 80,00%, maka hasil akurasi dan *F1 Score* tersebut dapat disimpulkan dengan hasil yang sangat baik dengan menggunakan *Algoritma Naïve Bayes* sehingga dapat dipakai untuk klasifikasi pada komentar Instagram.

IV. PENUTUP

A. Kesimpulan

Pada penelitian ini yang telah dilaksanakan, dapat mengambil kesimpulan bahwa:

1. Metode *Naïve Bayes* terbukti dapat melakukan analisis sentiment pada komentar Instagram secara otomatis oleh aplikasi *rapidminer*. Menggunakan data training dengan menggunakan data testing yang diambil pada komentar Instagram *theoriginote*, setiap komentar yang diperoleh diklasifikasikan dengan sentiment positif dan sentiment negative.
2. Hasil dari *performance* metode *Naïve Bayes* memiliki *accuracy* sebesar 78,48% nilai *class precision* 71,98% nilai *class recal* 93,30% dan nilai *F1 Score* 80,00%. Nilai tersebut sudah baik dalam melakukan klasifikasi dengan menggunakan algoritma *Naïve Bayes*.

B. Saran

Berdasarkan hasil analisis sentiment yang telah dilakukan dalam penelitian ini, penulis memberikan beberapa saran kepada peneliti yang akan melakukan analisis di masa yang akan datang di antaranya adalah:

1. Pada penelitian selanjutnya dapat menganalisis pada objek yang berbeda seperti menganalisis tentang capres 2024 yang akan datang.
2. Pada penelitian yang sama dapat mencoba menggunakan metode lainnya seperti Support Vector Machine (SVM) atau dengan metode lainnya untuk mengoptimalkan akurasi yang ada.
3. Penelitian selanjutnya juga dapat menganalisa dengan aplikasi yang berbeda seperti tiktok atau lemon8.
4. Penambahan huruf pada kata dengan menggunakan Bahasa Indonesia pada kamus gaul jaman sekarang dikarenakan pada kolom komentar Instagram terlalu banyak kata yang hurufnya double dan komentar dengan kata yang tidak baku.

V. REFERENCES

- [1] F. Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, p. 50, 2018, doi: 10.35314/isi.v3i1.335.
- [2] W. A. Prabowo and C. Wiguna, "Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 149, 2021, doi: 10.30865/mib.v5i1.2604.
- [3] H. Nurrnun Muchammad Shiddieqy, S. Paulus Insap, and W. Wing Wahyu, "Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen Di Twitter," *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2016, no. March, pp. 57–64, 2016.
- [4] E. W. Sandi Fajar Rodiansyah, "Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification," *Issues Contemp. Chinese Thought Cult.*, vol. 6, no. 1, pp. 3–33, 2012, doi: 10.1163/ej.9789004182127.i-302.6.
- [5] N. Ruhyana, "Analisis Sentimen Terhadap Penerapan Sistem Plat Nomor Gajil/Genap pada Twitter Dengan Metode Klasifikasi Naive Bayes," *J. IKRA-ITH Inform.*, vol. 3, no. 1, pp. 94–99, 2019, [Online]. Available: www.situs.com
- [6] E. Fadilah, "Implementasi Metode Profile Matching Terhadap Sistem Pendukung Keputusan Penerimaan Dana Zakat pada Badan Amil Zakat Pertamina (BAZMA)," *Matics*, vol. 10, no. 2, p. 39, 2019, doi: 10.18860/mat.v10i2.5745.
- [7] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *J. Intell. Syst. Comput.*, vol. 1, no. 1, pp. 43–49, 2019, doi: 10.52985/insyst.v1i1.36.