

ANALISIS SENTIMEN TERHADAP PANGAN NASIONAL PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES

Inda Sari Tomagola, Asep Id Hadiana, Puspita Nurul Sabrina

Teknik Informatika, Universitas Jenderal Achmad Yani
Jl. Jenderal Achmad Yani, Cibeer, Kec. Cimahi Sel., Kota Cimahi,
Indasaritomagola146@gmail.com

ABSTRAK

Dataset di dapatkan dengan cara *crawling* data pada twitter setelah dataset dikumpulkan, kemudian dilakukan *preprocessing* untuk mengoptimalkan pengolahan data. *Preprocessing* terdiri dari empat tahap, yang mencakup, *case folding*, *tokenizing*, *filtering* dan *stemming*. Pembobotan kata memberikan nilai pada kata-kata yang ada di dokumen. berbeda-beda tergantung pada metode yang digunakan. Sesuai dengan algoritma yang digunakan pada penelitian ini menggunakan algoritma pembobotan kata TF-IDF dan TF-RF. Pengujiannya akan dilakukan dengan menggunakan metode klasifikasi naïve Bayes dan analisis perbandingannya dengan *confusion matrix*. Dengan dilakukan beberapa eksperimen 80:20 memberikan tingkat akurasi terbaik dari dua eksperimen lainnya dengan hasil akurasi TF-IDF naïve bayes 73% dan TF-RF naïve 72%. Sehingga dari kedua metode pembobotan tersebut dapat disimpulkan TF-IDF naïve bayes lebih baik hasil akurasi dibandingkan hasil TF-RF naïve bayes.

Kata kunci: Twitter, text mining, preprocessing, feature extraction, tf-idf, tf-rf, naive bayes,

1. PENDAHULUAN

Proses mengubah data masukan menjadi kumpulan fitur untuk mendapatkan representasi yang paling sedikit dari data masukan dikenal sebagai ekstraksi fitur. [1]. Feature extraction adalah salah satu teknik untuk mengenali objek berdasarkan histogramnya[2]. Ada beberapa algoritma yang digunakan dalam proses text mining untuk memberikan nilai atau berat pada setiap kata dalam dokumen. Metode yang digunakan menentukan berat atau nilai yang diberikan pada setiap kata. Misalnya, beberapa algoritma pembobotan kata digunakan, seperti TF, Idf, RF, TF-IDF, TF.RF, dan WIDF.[3].

Dengan feature extraction, kata dihubungkan ke vektor numerik yang dapat digunakan dalam komputasi. One-hot encoding, yang menggunakan panjang vektor untuk menampilkan satu kata, adalah metode paling sederhana untuk mengkarakterisasi teks. Posisi kata ditentukan oleh teks atau kata-kata di sekitarnya. Dengan mengembedding kata, makna sintaktik dan semantiknya dapat ditangkap. Penelitian ini menggunakan model ekstraksi fitur TF-IDF dan TF-RF, yang dapat memberikan nilai atau bobot pada term di dokumen algoritma pengajaran mesin untuk klasifikasi teks multilabel, yaitu naïve bayes. Tujuan dari penelitian ini adalah untuk membandingkan kedua metode pembobotan yang akan dihitung secara terpisah untuk menentukan kombinasi pembobotan kata yang paling akurat.[4]

Banyak platform media sosial saat ini memungkinkan orang untuk mengekspresikan pendapat, pendapat, dan kritik. Setiap hari, banyak orang memposting di media sosial, salah satunya Twitter. Opini ini mencakup pendapat negatif dan positif tentang masalah atau peristiwa yang terjadi di masyarakat. [5]. Opini publik yang terjadi di media

sosial dapat digunakan untuk menganalisis sentimen tentang ketahanan pangan. [6].

Dengan penelitian ini bagaimana menganalisis klasifikasi sentimen dari tingkat akurasi analisis sentiment twitter tentang pangan dengan menggunakan tiga kelas label positif, negatif dan netral dengan menggunakan teknik SMOTE untuk menyeimbangkan kelas label agar dapat menghasilkan nilai akurasi yang baik dari penelitian sebelumnya untuk feature extraction yang digunakan adalah TF-IDF dan TF-RF, dengan cara membandingkan kedua metode tersebut menggunakan klasifikasi naïve bayes apakah hasil akurasi dari TF-IDf naïve bayes lebih baik atau hasil akurasi antara TF-Rf naïve bayes.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentiment adalah cabang yang menganalisis komentar individu dan sentimen publik[7]. Analisis sentimen secara otomatis memproses data untuk mengumpulkan sentimen dari kalimat; ini dilakukan untuk menentukan apakah opini publik positif atau negatif[6]. Sebagian besar model analisis sentimen saat ini menggunakan kata-kata opini ini sebagai indikator penting dari pendapat publik. Beberapa kamus opini publik dalam literatur termasuk SentiWordNet, General Inquirer, dan SenticNet.[7].

Analisis sentimen adalah teknik yang secara otomatis mengekstraksi, mengolah, dan memahami teks yang tidak terstruktur untuk mendapatkan informasi tentang pendapat atau opini yang terkandung dalam kalimat.[8].

2.2. Text Mining

Text mining juga dikenal sebagai data mining teks atau pengambilan informasi dari basis data teks. Tujuan dari text mining adalah untuk menemukan

informasi atau mendapatkan inti informasi yang kemudian dapat diolah untuk diproses.[9]. Text mining juga berarti menggali di komputer untuk menemukan hal-hal baru atau informasi yang mudah dipahami. Hasilnya didasarkan pada data yang diambil secara otomatis dari berbagai sumber data teks.[3].

2.3. Lexicon Based Features (sentiment scoring)

Metode *lexicon-based* dengan terlebih dahulu membuat kamus kata opini (*lexicon*). Kata-kata dalam kamus dapat digunakan untuk mengetahui apakah suatu kalimat yang mengandung opini atau tidak[10]. Pada *lexicon* memiliki sentiment positif atau negatif yang diberi bobot berdasarkan kamus *lexicon*. Tujuan penggunaan fitur Lexicon-Based adalah untuk menentukan orientasi sentimen dari sebuah kata [11]. *Lexicon-based* bertujuan untuk mengambil informasi dari media sosial twitter [12].

2.4. Term Frequency-Inverse Document Frequency (TF-IDF)

Metode *Term Frequency Invers Document Frequency* (TF-IDF) adalah metode yang menentukan seberapa sering sebuah kata (term) terhubung ke dokumen dan memberikan nilai pada setiap kata. [13]. Tf-idf adalah salah satu cara untuk mengatasi keterbatasan ini adalah dengan menekankan hubungan kata (term) dalam dokumen. [14]. Frekuensi kata yang muncul dalam dokumen menunjukkan pentingnya kata tersebut. Jumlah dokumen yang mengandung kata tersebut dan frekuensi di mana kata tersebut muncul. Oleh karena itu, bobot hubungan antara kata dan dokumen tinggi jika frekuensi kata dalam dokumen tinggi dan total frekuensi kata dalam dokumen rendah[15],[16].

2.5. Term Frequency -Relevance Frequency (TF-RF)

TF-RF berasal dari TF-IDF, namun kelemahan TF-IDF adalah teknik ini tidak memiliki kemampuan untuk membedakan dokumen yang positif dan negatif? [17]. *Relevance Frequency* dikenal juga sebagai RF, adalah sebuah metode yang relatif baru, munculnya metode ini merupakan upaya untuk menyempurnakan metode yang ada. Metode ini mempertimbangkan pentingnya dokumen yang dinyatakan dengan frekuensi istilah dalam kategori terkait RF menilai relevansi dokumen untuk setiap kata dalam kelas tertentu; kata-kata dinilai berdasarkan frekuensi relevansinya[18].

Term Frequency-Relevance Frequency (TF-RF) kemunculan metode ini diharapkan untuk meningkatkan metode-metode yang sudah ada, metode ini akan menilai istilah berdasarkan kemunculan (ada atau tidaknya) dalam sebuah dokumen. Metode RF, yang pertama kali diusulkan oleh Man Lan, memungkinkan untuk mempertimbangkan pentingnya dokumen terkait frekuensi istilah dalam kategori yang relevan[19].

2.6. Naïve Bayes

Naive Bayes merupakan algoritma klasifikasi yang didasarkan pada teorema Bayes dan adjective naive didasarkan pada asumsi bahwa karakteristik data yang tidak bergantung satu sama lain[20]. Teknik klasifikasi Naive Bayes merupakan metode lama yang masih sangat populer dalam proses analisis sentimen[14]. Ada banyak penelitian yang menggunakan algoritma naive bayes untuk kalsifikasi sentimen. Keunggulan naive bayes adalah tidak membutuhkan data pelatihan yang besar untuk menentukan estimasi parameter yang dibutuhkan dalam proses klasifikasi.[21].

2.7. Confusion Matrix

Confusion matrix adalah metrik kinerja untuk masalah klasifikasi mechine learning yang dapat menampilkan dua kelas atau lebih[22]. Confusion matrix merupakan tabel yang berisi 4 kombinasi berbeda dari nilai prediksi dan nilai aktualnya.

Terdapat empat istilah yang merupakan representasi hasil dari kalsifikasi confusion matrix contoh pada gambar 2.1 dan gambar 2.2 dimana TP (True Positif) merupakan data positif yang di prediksi benar, TN (True Negatif) menyatakan data yang negatif di prediksi benar kemudian pada FP (False Positif) sebagai kesalahan tipe 1 merupakan data negatif namun di prediksi sebagai data positif, sebaliknya dengan FN (False Negatif) sebagai kesalahan tipe 2 merupakan data positif yang di prediksi sebagai data negative, selain itu untuk menghitung nilai tersebut di antaranya Accuracy, Precision, Reccal dan F1-Score [22]. Confusion matrix adalah metode penilaian yang menggunakan tabel matriks yang ditunjukkan di bawah ini:

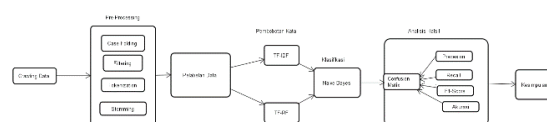
		Prediksi	
		TRUE	FALSE
Aktual	TRUE	TP	FP
	FALSE	FN	TN

Gambar 1. Confusion Matrix

3. METODE PENELITIAN

3.1. Tahapan Penelitian

Pada penelitian ini ada beberapa tahapan yang dilakukan mulai dari perolehan data, preprocessing, pembobotan kata, klasifikasi, analisis hasil, kesimpulan.



Gambar 2. Tahapan Penelitian

Berdasarkan gambar diatas mengenai tahapan penelitian pada gambar 1. Berikut ini merupakan penjelasannya:

1. Perolehan Data

Data yang di peroleh dari penelitian dengan cara *crawling* data tweet yang menggunakan kata kunci “pangan”. Pada teknik *crawling* ini dengan bantuan aplikasi node JS dan command prompt.

2. Preprocessing

Pre-processing merupakan langkah-langkah dalam penelitian yang memperbaiki data sehingga lebih terstruktur untuk diproses di setiap dokumen yang digunakan[18]. Langkah-langkah yang terlibat dalam preprocessing teks meliputi berikut ini:

a) *Case Folding*

Case folding merupakan tahapan dimana mengubah semua huruf besar dalam kalimat menjadi huruf kecil. Hanya huruf ‘a’ sampai dengan ‘z’ yang diterima agar sistem dapat mengolah data yang lebih efisien dan efektif.

b) *Filtering*

Filtering merupakan suatu proses dimana mengambil kata-kata penting dari sebagian data tertentu, dan menghilangkan karakter, tanda baca serta kata-kata yang tidak memiliki makna atau informasi yang dibutuhkan.

c) *Tokenizing*

Tokenization merupakan proses pemotongan string masukan berdasarkan tiap kata dari sebuah dokumen menjadi sebuah potongan-potongan kata kata atau karakter sesuai dengan kebutuhan sistem. *Tokenization* dilakukan untuk mempermudah dalam menentukan setiap kata tersebut termasuk kedalam sentimen positif, negatif atau netral.

d) *Stemming*

Stemming merupakan proses dimana mengubah token yang asalnya memiliki imbuhan menjadi kata dasar dan menghilangkan imbuhan.

3. Feature Extraction

Setelah pre-processing data untuk tahap selanjutnya yaitu feature extraction. Pada penelitian ini feature extraction yang digunakan yaitu TF-IDF dan TF-RF.

- a. TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah algoritma pembobotan yang memberikan bobot pada hubungan antara kata dan dokumen.

$$tf_{td}idf_t = tf_{td} * \log\left(\frac{N}{df_t}\right)$$

Dimana:

*tf * idf*: adalah bobot total dari term t

tftd: adalah jumlah kemunculan term t dalam suatu dokumen

N: adalah total dari keseluruhan dokumen

dft: adalah jumlah dari keseluruhan dokumen yang mengandung term

- b. TF-RF (*Term Frequency-Relevance Frequency*) adalah algoritma yang hanya akan menilai istilah berdasarkan kemunculan (ada atau tidaknya) dalam sebuah dokumen.

$$tf_{td}rf = tf_{td} * \log\left(2 + \frac{b}{\max(1, c)}\right)$$

Dimana:

*tf * rf*: merupakan bobot dokumen dalam model ruang vektor

tftd: menunjukkan berapa kali kata t muncul dalam sebuah dokumen

b: menunjukkan jumlah dokumen yang mengandung kata t tersebut

c: menunjukkan jumlah dokumen yang tidak mengandung kata t tersebut

4. Klasifikasi Naive Bayes

Setelah melakukan pembobotan kata sehingga mendapatkan nilai bobot pada tahapan selanjutnya yaitu pengklasifikasian menggunakan algoritma naive bayes dalam penelitian ini yang nantinya memberikan nilai akurasi. Apakah nilai akurasi pada TF-IDF naive bayes lebih baik atau TF-RF naive bayes. Naive Bayes

Naive Bayes merupakan algoritma klasifikasi yang didasarkan pada teorema Bayes dan adjective naive didasarkan pada asumsi bahwa karakteristik data yang tidak bergantung satu sama lain[20]. Teknik klasifikasi Naive Bayes merupakan metode lama yang masih sangat populer dalam proses analisis sentimen[14] ada banyak penelitian yang menggunakan algoritma naive bayes untuk kalsifikasi sentimen. Keunggulan naive bayes adalah tidak membutuhkan data pelatihan yang besar untuk menentukan estimasi parameter yang dibutuhkan dalam proses klasifikasi.[21]. Pada peneltian sebelumnya yang menggunakan naive bayes sebagai metode klasifikasi, dimana secara keseluruhan Penggunaan metode Naive Bayes memberikan hasil yang baik klasifikasi dalam tingkat akurasi 92,2% [20].

$$P(H|X) = \frac{P(P|H) P(H)}{P(X)}$$

Dimana:

P(H|X): Adalah probabilitas kelas H dari dokumen input (tweet), probabilitas posterior

P(X|H): Adalah probabilitas suku X di kelas H, probabilitas bersyarat

P(H): Apakah probabilitas kelas H terjadi dalam dataset, probabilitas kelas sebelumnya

P(X): Apakah probabilitas istilah X terjadi dalam data probabilitas

Saat menghitung prediksi sebelumnya, kemungkinan mengklasifikasikan kemunculan kata yang sebenarnya dapat dihilangkan, karena kemungkinan tersebut tidak mempengaruhi perbandingan hasil klasifikasi untuk setiap kategori. Oleh karena itu, proses klasifikasi dapat disederhanakan dengan menggunakan persamaan berikut [23].

Menghasilkan data nol untuk kelas ini. Oleh karena itu, metode pemulusan Laplace digunakan untuk mencegah situasi probabilitas nol dan memastikan bahwa setiap kata memiliki peluang untuk muncul berdasarkan kemunculannya, setidaknya satu besaran dalam persamaan berikut:

$$P(H|X) = P(H_i) \prod_{k=1}^n P(x_k|H_i)$$

5. Evaluasi Hasil

Hasil evaluasi dapat mengklasifikasikan sentimen dalam kategori positif, negatif, atau netral. Dengan hasil perbedaan antara algoritma pembobotan TF-IDF dan TF-RF menggunakan algoritma klasifikasi naïve bayes.

$$Precision = \frac{TP}{TP + FP}$$

Precision : Ini adalah rasio jumlah sentimen yang di prediksi secara akurat dengan jumlah total sentimen yang di prediksi[24].

$$Recall = \frac{TP}{TP + FN}$$

Recall : Merupakan rasio jumlah sentiment yang di prediksi secara akurat dengan jumlah sentiment aktual[24].

$$F - 1 \text{ Score} = \frac{2 * (Recall * Precision)}{Recall + Precision}$$

F-1 Score : Ini adalah hasil dari perbandingan dari hasil rata-rata precision dan hasil recall. Accuracy dapat digunakan sebagai tolok ukur kinerja algoritma ketika dataset memiliki jumlah false negatif dan false positif yang sangat mendekati (symmetric). Namun, jika jumlahnya tidak mendekati, lebih baik menggunakan F-1 Score sebagai pembandingan[24].

$$Accuracy = \frac{TP \text{ tiap kelas}}{Total \text{ data}}$$

Accuracy 3 kelas : Mengukur tingkat akurasi dari kalsifikasi data yang akan dievaluasi pada 3 kelas [24].

6. Kesimpulan

Pada tahapan ini dimana membuat kesimpulan dari hasil evaluasi yang didapatkan pada penelitian ini,

$$P(H_i) \times P(x_1|H_i) \times P(x_2|H_i) \dots \times P(x_n|H_i)$$

4. HASIL DAN PEMBAHASAN

Berbagai jenis data digunakan dalam penelitian ini, termasuk data training dan data testing. Data traning digunakan untuk membuat model klasifikasi, yang merupakan representasi pengetahuan yang akan digunakan untuk memprediksi dokumen data baru yang belum pernah ada. Proses mining data ini akan mencakup 1632 dokumen dokumen yang telah dikelompokkan berdasarkan sentimen yang terkait. Data ini memiliki tiga sentiment yaitu: negatif, netral, dan positif.

Tabel 1. Pelabelan Data

No	Tweet	Sentimen
1	benar bicara kena lapar bukan swasembada pangan kebiasaan campur adu	Negatif
2	Salah satu syarat wujud swasembada pangan Sejahtera tani	Netral

No	Tweet	Sentimen
3	anda subsidi kendara listrik alih subsidi pupuk pangan lebih beri impact besar masyarakat	Positif
4	bagi prestasi anies baswedan lama pimpin jakarta hadirkan pangan murah npenataan transportasi demokratis gak penjara rakyat kritik bongkarkemenkeugate bongkar kemenkeugate	Netral
6	lama gaung swasembada pangan sejahtera tani n nsayang sekali capres maupun pimpin negeri punya program sebut periode padahal sederhana	Positif
...		
1632	coba berfikir sadar qt apa alami negri dulu bangga pabrik pupuk sriwijaya sekarang kiprahx sadar tahan pangan mati hingga qt negara besar paksa import butuh pokok	Positif

4.1. Klasifikasi TF-IDF Naïve Bayes

Jika telah mendapatkan hasil pembobotan dari TF-IDF selanjutnya melakukan tahap kalsifikasi menggunakan naïve bayes. Untuk mendapatkan nilai probabilitas yang tinggi dilakukan tiga tahapan sesuai dengan data latih, data latih yang digunakan untuk uji cob aini menggunakan data latih yang sebelumnya dapat dilihat pada tabel berikut.

Tabel 2. Data Latih Klasifikasi TF-IDF dan TF-RF

Fitur	Teks			Label
	D1	D2	D3	
benar	0.477	0.176	0	negatif
swesembada	0	0.176	0	netral
...
masyarakat	0	0	0.477	postif

4.1.1. Hasil Eksperimen 80:20

Untuk eksperimen 80:20 data yang dibagi 80% untuk data latih dan 20% untuk data uji, dan dari hasil eksperimen 80:20 dilihat bahwa terdapat perbedaan antara TF-IDF+NB dan TF-RF+NB. Setelah mendapatkan nilai dari TP, TN, FP, FN diatas selanjutnya dapat dihitung nilai *precesion*, *Recall*, *F1-Score* dari tiap kelasnya serta nilai akurasinya hasilnya dapat dilihat pada tabel 3. untuk TF-IDF dan 4 untuk TF-RF

Tabel 3. TF-IDF

	TP	FP	TN	FN
Negatif	131	56	238	13
Positif	78	25	265	70
Netral	113	35	257	33

Tabel 4. TF-RF

	TP	FP	TN	FN
Negatif	128	62	232	16
Positif	77	31	259	71
Netral	112	28	232	34

Tabel 5. Perhitungan Precision, Recall, F-1Score

Label	Akurasi	Precision	Recall	F1 Score
Negatif	0,89	0,67	0,89	0,77
Netral	0,52	0,71	0,52	0,60
Positif	0,77	0,80	0,77	0,78
rata-rata	0,74	0,74	0,74	0,72

Tabel 6. Akurasi TF-RF

	Precesion (%)	Recall (%)	F1-Score (%)
Negatif	67%	89%	77%
Positif	80%	77%	78%
Netral	71%	52%	60%
Accuracy (%)	72%		

Kelas Negatif

$$\text{Precesion} : \frac{TP}{TP+FP} = \frac{128}{128+62} = 0.67$$

$$\text{Recall} : \frac{TP}{TP+FN} = \frac{128}{128+16} = 0.89$$

$$\text{F1-Score} : \frac{2 \times (\text{Recall} \times \text{Precesion})}{\text{Recall} + \text{Precesion}} = \frac{2 \times (0.89 \times 0.67)}{0.89 + 0.67} = \frac{1.2}{1.56} = 0.77$$

$$\text{Akurasi} : \frac{Tp \text{ tiap kelas}}{\text{Total data}} = \frac{128+77+112}{438} = 0.72$$

Tabel 7. Perhitungan Precision, Recall, F-1Score dan Akurasi TF-IDF

Label	Akurasi	Precision	Recall	F1 Score
Negatif	0,91	0,70	0,91	0,79
Netral	0,53	0,76	0,53	0,62
Positif	0,77	0,76	0,77	0,77
rata-rata	0,75	0,75	0,75	0,74

Kelas Negatif

$$\text{Precesion} : \frac{TP}{TP+FP} = \frac{131}{131+56} = 0.70$$

$$\text{Recall} : \frac{TP}{TP+FN} = \frac{131}{131+13} = 0.91$$

$$\text{F1-Score} : \frac{2 \times (\text{Recall} \times \text{Precesion})}{\text{Recall} + \text{Precesion}} = \frac{1.27}{1.61} = 0.79$$

$$\text{Akurasi} : \frac{Tp \text{ tiap kelas}}{\text{Total data}} = \frac{132+178+113}{438} = 0.73$$

Tabel 8 Akurasi TF-IDF

	Precesion (%)	Recall (%)	F1-Score (%)
Negatif	70 %	91%	79%
Positif	77%	76%	77%
Netral	76%	53%	62%
Accuracy (%)	73%		

4.2. Implementasi Sistem

Implementasi sistem menampilkan halaman Beranda, Input Dataset, Klasifikasi.

4.2.1. Implentasi Halaman Home

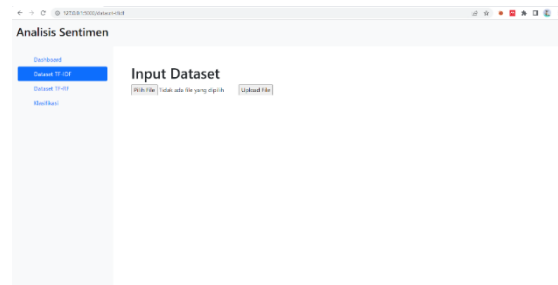
Implementasi halaman home merupakan tampilan yang terlihat Ketika masuk pertama kali pada sistem yang dibangun dengan menampilkan judul, nama pembuat hingga nim. Implementasi dari halaman home dapat dilihat pada gambar dibawah ini.



Gambar 3. Implementasi Halaman Home

4.2.2. Implementasi Halaman Input Dataset

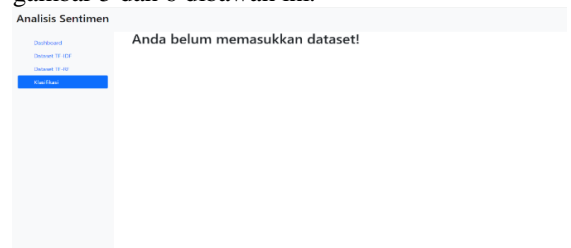
Implementasi halaman dataset merupakan halaman yang menampilkan dataset yang digunakan dalam penelitian ini, dan dataset yang ditampilkan sudah ada labelnya. Halaman dataset dapat dilihat pada gambar 4.



Gambar 4. Input Dataset

4.2.3. Implentasi Halaman Klasifikasi

Implementasi halaman klasifikasi merupakan halaman yang menampilkan hasil kalsifikasi dari penelitian ini, pada proses kalsifikasi untuk membandingkan hasil akurasi dari metode TF-IDF dan TF-RF dengan menggunakan naïve bayes. Implementasi halaman kalsifikasi dapat dilihat pada gambar 5 dan 6 dibawah ini.



Gambar 5. Implentasi Halaman belum ada proses Klasifikasi



Gambar 6. Implentasi Halaman setelah ada proses Klasifikasi

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian ini tingkat keakurasian dengan menggunakan analisis sentimen twitter tentang pangan dengan menggunakan kinerja pembobotan TF-IDF dan TF-RF terhadap proses klasifikasi menggunakan naïve bayes dengan data yang digunakan sebesar dataset yang diambil menggunakan teknik *crawling* dengan menggunakan node JS sebagai aplikasi pendukung. Dari hasil pengujian diatas dapat disimpulkan bahwa tingkat akurasi antara TF-IDF naïve bayes dan TF-RF naïve bayes dengan beberapa eksperimen dengan 80:20 hasil akurasi untuk TF-IDF naïve bayes 73% kemudian nilai akurasi untuk TF-RF naïve bayes dengan hasil akurasi 72%, hasil eksperimen 70:30 nilai akurasinya 72% untuk TF-IDF naïve bayes, untuk TF-RF naïve bayes 71% dan hasil eksperimen 60:40 hasil akurasi TF-IDF 69% TF-RF 69% Sehingga dari kedua metode pembobotan tersebut dapat disimpulkan TF-IDF naïve bayes lebih baik hasil akurasi dibandingkan hasil TF-RF naïve bayes dengan melakukan eksperimen 80:20. Berdasarkan hasil penelitian ini, ada beberapa saran untuk pengembangan lebih lanjut: Dari hasil yang sudah penulis dapatkan untuk ekstraksi fitur pada pembobotan kata menggunakan teknik selain TF-IDF dan TF-RF untuk mendapatkan hasil yang lebih akurat. Mungkin pada penelitian selanjutnya dapat menggunakan teknik pembobotan kata dalam *text processing* dan juga dapat melakukan pengujian menggunakan teknik klasifikasi lainnya.

DAFTAR PUSTAKA

- [1] K. Jasa, "Bab 2 Landasan Teori," *Apl. dan Anal. Lit. Fasilkom UI*, vol. m, no. 1998, pp. 7–34, 2000, [Online]. Available: http://elib.unikom.ac.id/files/disk1/655/jbptunik_ompp-gdl-supriadini-32740-6-12.unik-i.pdf.
- [2] Y. Religia, "Feature Extraction Untuk Klasifikasi Pengenalan Wajah Menggunakan Support Vector Machine Dan K-Nearest Neighbor," *Pelita Teknol. J. Ilm. Inform. Arsit. dan Lingkung.*, vol. 14, no. 2, pp. 85–92, 2019.
- [3] F. V. Sari and A. Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," *J. SIMETRIS*, vol. 10, no. 2, pp. 681–686, 2019.
- [4] L. Efrizoni, S. Defit, M. Tajuddin, and A. Anggrawan, "Komparasi Ekstraksi Fitur dalam Klasifikasi Teks Multilabel Menggunakan Algoritma Machine Learning," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 3, pp. 653–666, 2022, doi: 10.30812/matrik.v21i3.1851.
- [5] A. Tirtayasa and A. Listiyo Wibowo, "Sentiment Analysis Tweet KTT G-20 di Media Sosial Twitter Menggunakan Metode Naïve Bayes," *J. Pengemb. Sist. Inf. dan Inform.*, vol. 4, no. 2, pp. 1–12, 2023, doi: 10.47747/jpsii.v4i2.1097.
- [6] N. M. A. J. Astari, Dewa Gede Hendra Divayana, and Gede Indrawan, "Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier," *J. Sist. dan Inform.*, vol. 15, no. 1, pp. 27–29, 2020, doi: 10.30864/jsi.v15i1.332.
- [7] B. Agarwal, N. Mittal, P. Bansal, and S. Garg, "Sentiment analysis using common-sense and context information," *Comput. Intell. Neurosci.*, vol. 2015, 2015, doi: 10.1155/2015/715730.
- [8] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.
- [9] A. Rahayu, "Bab iii landasan teori 3.1.," *http://e-journal.uajy.ac.id/7244/4/3TF03686.pdf*, no. 2010, pp. 15–48, 2019, [Online]. Available: <http://e-journal.uajy.ac.id/7244/4/3TF03686.pdf>.
- [10] Y. Azhar, "Metode Lexicon-Learning Based Untuk Identifikasi Tweet Opini Berbahasa Indonesia," *J. Nas. Pendidik. Tek. Inform.*, vol. 6, no. 3, p. 237, 2018, doi: 10.23887/janapati.v6i3.11739.
- [11] R. Catelli, S. Pelosi, and M. Esposito, "Lexicon-Based vs. Bert-Based Sentiment Analysis: A Comparative Study in Italian," *Electron.*, vol. 11, no. 3, 2022, doi: 10.3390/electronics11030374.
- [12] M. Metode, F. Multiple, C. Decision, M. Fmcdm, and D. Yogyakarta, "Indonesian Journal of Business Intelligence," vol. 3, no. 2, pp. 54–60, 2020.
- [13] A. Deolika, K. Kusriani, and E. T. Luthfi, "Analisis Pembobotan Kata Pada Klasifikasi Text Mining," *J. Teknol. Inf.*, vol. 3, no. 2, p. 179, 2019, doi: 10.36294/jurti.v3i2.1077.
- [14] Merinda Lestandy, Abdurrahim Abdurrahim, and Lailis Syafa'ah, "Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 802–808, 2021, doi: 10.29207/resti.v5i4.3308.
- [15] P. T. Dan and R. Logistik, "Analisis Sentimen Twitter tentang Covid-19 Menggunakan Istilah," pp. 14–16, 2021.
- [16] S. Ghosh and M. S. Desarkar, "Class Specific TF-IDF Boosting for Short-text Classification," pp. 1629–1637, 2018, doi: 10.1145/3184558.3191621.
- [17] I. Istiqlaliyyah, Y. A. Sari, and M. A. Fauzi, "Pencarian Teks pada Terjemahan Ayat Al- Qur'an dengan Menggunakan TF-RF dan Bray-Curtis Distance," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 5, pp. 4357–4363, 2019.
- [18] A. N. Assidyk, E. B. Setiawan, and I. Kurniawan, "Analisis Perbandingan Pembobotan TF-IDF dan TF-RF pada Trending Topic di Twitter dengan Menggunakan Klasifikasi K-Nearest Neighbor," *e-Proceeding Eng.*, vol. 7, no. 2, pp. 7773–7781,

- 2020.
- [19] M. Syulhan, A. Ghofany, R. D. B, and F. Bimantoro, *Proceedings of the First Mandalika International Multi-Conference on Science and Engineering 2022, MIMSE 2022 (Informatics and Computer Science)*, vol. 1, no. 1. Atlantis Press International BV, 2022.
- [20] A. Muzaki and A. Witanti, "Sentiment Analysis of the Community in the Twitter To the 2020 Election in Pandemic Covid-19 By Method Naive Bayes Classifier," *J. Tek. Inform.*, vol. 2, no. 2, pp. 101–107, 2021, doi: 10.20884/1.jutif.2021.2.2.51.
- [21] F. Ilzam Nur Haq and E. Budi, "Implementasi Naive Bayes Classifier untuk Prediksi Kepribadian Big Five pada Twitter Menggunakan Term Frequency-Inverse Document Frequency (TF-IDF) dan Term Frequency-Relevance Frequency (TF-RF) Program Studi Sarjana Ilmu Komputasi Fakultas Informatik," *e-Proceeding Eng.*, vol. 6, no. 2, pp. 9785–9795, 2019.
- [22] Prameswari Reksa Agami, I. Yuniar Purbasari, and B. Rahmat, "Penentuan Penggunaan Lulur Dan Masker Organik Sesuai Dengan Diagnosa Jenis Kulit Wajah Menggunakan Metode Decision Tree Algoritma C4.5," *J. Inform. dan Sist. Inf.*, vol. 2, no. 2, pp. 313–321, 2021, doi: 10.33005/jifosi.v2i2.351.
- [23] P. Antinasari, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 12, pp. 1718–1724, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [24] D. Normawati and S. A. Prayogi, "Implementasi Naive Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021, [Online]. Available: <http://ejournal.tunasbangsa.ac.id/index.php/jsakti/article/view/369>.