

ANALISIS SENTIMEN MENGGUNAKAN NBC DAN SVM UNTUK MENGETAHUI RESPON TERHADAP KINERJA PRESIDEN MELALUI MEDIA INSTAGRAM

Amir Hamzah^{1*}, Renna Yanwastika Ariyana², Uning Lestari³
^{1,2,3} Informatika Universitas Akprind Indonesia, *Penulis Koresponden
e-mail:¹amir@akprind.ac.id,²renna@akprind.ac.id,³uning@akprind.ac.id,

ABSTRACT

The president's performance can be revealed through the public's response to various presidential programs in the economic or industrial sector which are uploaded on the President's official Instagram media. The public's response to each uploaded post is interesting data to analyze. This research aims to conduct sentiment analysis of the public's response to the president's performance, by uploading the president's activities on Instagram using the Naive Bayes Classifier (NBC) and Support Vector Machine (SVM) methods. A total of 18,660 responses to the President's 5 Instagram posts on economic and industrial activities between March and September 2024. The data was then filtered to just 15,306 comments which were determined as a dataset. Dataset labeling before classification is carried out using the lexicon based method. The research results show that sentiment towards the president's performance is 28.3% positive, 36.8% neutral and 34.9% negative. Meanwhile, the NBC and SVM classification results obtained showed an accuracy of 68.7%, precision of 68.5%, recall of 67.8% and F1-score of 68.2% for the NBC algorithm. The SVM algorithm produces an accuracy of 80.5%, a precision value of 80.5%, a recall value of 79.6%, and an F1-score value of 79.6%. It can be seen that the SVM algorithm is superior to the NBC algorithm in all classification parameters. From several training and testing compositions, it was also found that the best composition was the 70:30 composition.

Keywords: classification, presidential performance, NBC, SVM

INTISARI

Kinerja presiden dapat diungkap melalui respon masyarakat terhadap berbagai program presiden di bidang ekonomi atau industri yang diunggah di media Instagram resmi milik Presiden. Respon masyarakat pada setiap posting yang diunggah, merupakan data yang menarik untuk dianalisis. Penelitian ini bertujuan melakukan analisis sentimen terhadap respon masyarakat pada kinerja presiden yang, melalui unggahan kegiatan presiden di Instagram menggunakan metode *Naive Bayes Classifier* (NBC) dan *Support Vector machine* (SVM). Sebanyak 18.660 respon terhadap 5 postingan Instagram Presiden pada kegiatan bidang ekonomi dan industri antara rentang waktu bulan Maret sampai September 2024. Data tersebut kemudian dilakukan filtering menjadi 15.306 komentar saja yang ditetapkan sebagai dataset. Pelabelan dataset sebelum dilakukan klasifikasi dilakukan dengan menggunakan metode *lexicon based*. Hasil penelitian menunjukkan bahwa sentimen terhadap kinerja presiden adalah 28.3% positive, 36.8% netral dan 34.9% negatif. Sedangkan hasil klasifikasi NBC dan SVM diperoleh menunjukkan *accuration* sebesar 68.7%, *precision* sebesar 68.5%, *recall* sebesar 67.8% dan *F1-score* 68.2% untuk algoritma NBC. Untuk algoritma SVM menghasilkan *accuration* sebesar 80.5%, nilai *precision* 80.5%, nilai *recall* 79.6%, dan nilai *F1-score* 79.6%. Terlihat bahwa algoritma SVM lebih unggul dari pada algoritma NBC pada semua parameter klasifikasi. Dari beberapa komposisi training dan testing juga didapatkan komposisi terbaik adalah komposisi 70:30.

Kata kunci: klasifikasi, kinerja presiden, NBC,SVM

1. PENDAHULUAN

Menjelang akhir masa jabatan Presiden Jokowi, setelah dua periode masa jabatan, banyak survei dilakukan oleh lembaga-lembaga survei tentang tingkat kepuasan masyarakat terhadap kinerja Presiden. Beberapa survei memberikan nilai kepuasan masyarakat yang cukup tinggi, misalnya survei yang dilakukan oleh LSI Denny JA pada tanggal 26 September - 3 Oktober 2024 dengan 1200 responden yang memberikan nilai kepuasan sebesar 80,8% dari 1200 responden (Nufus, 2024). Survei lain yang dilakukan oleh Lembaga Indikator politik tanggal 22-29 September dengan responden sebanyak 3.540 orang juga memberikan kepuasan cukup tinggi, yaitu sebesar

75% (Chaterine dan Prabowo, 2024). Metode survei yang dilakukan hampir sama, yaitu dengan memberikan kuesener dengan beberapa pertanyaan tentang kinerja presiden selama menjabat. Namun meskipun hasil survey menunjukkan kepuasan masyarakat yang cukup tinggi, menurut Mukhroaman et.al (2024) hasil kepuasan masyarakat sebagaimana ditunjukkan oleh survei belum tentu menunjukkan kinerja yang sebenarnya. Hal ini karena bisa jadi penggunaan survei memang dirancang untuk mempengaruhi opini masyarakat. Untuk itu diperlukan opini alternatif dari pandangan masyarakat yang lebih bebas, yakni dengan menangkap opini masyarakat pada portal resmi presiden. Opini masyarakat ini dapat dianalisis dari respon masyarakat terhadap berbagai berita yang diposting dalam portal menggunakan analisis sentimen.

Analisis sentimen saat ini merupakan alat analisis yang terus berkembang seiring dengan perkembangan big data teks yang melonjak eksponensial (Mao,et.al., 2024). Perkembangan teknologi komunikasi dan informasi yang sangat cepat dapat dimanfaatkan oleh para pelaksana kebijakan pemerintah untuk mendukung kinerja dan menyerap aspirasi masyarakat. Presiden sebagai pusat pemerintahan telah memanfaatkan berbagai aplikasi media sosial untuk mendukung kinerja presiden dengan membangun komunikasi publik dan menyerap aspirasi masyarakat melalui media tersebut (Mandala, 2024). Banyak kalangan bahkan menilai bahwa media tersebut bukan saja untuk mengkomunikasikan kebijakan, tetapi yang lebih menonjol adalah untuk membangun pencitraan (Fadiyah dan Simorangkir,2019). Saat ini salah satu media sosial Presiden adalah Instagram dengan akun resmi <https://www.instagram.com/jokowi/>. Akun tersebut memiliki *follower* 58.9 juta dengan 5.677 post yang berisi berbagai kegiatan presiden, termasuk kegiatan yang terkait dengan kegiatan dalam pembangunan ekonomi. Komentar Masyarakat untuk berbagai kegiatan tersebut merupakan data yang menarik karena dapat digunakan untuk menilai sentiment dan respon masyarakat terhadap kinerja presiden pada bidang tertentu. Untuk itu penelitian ini memilih media Instagram presiden untuk menilai kinerja presiden berdasarkan respon masyarakat pada berbagai posting yang terkait dengan bidang ekonomi.

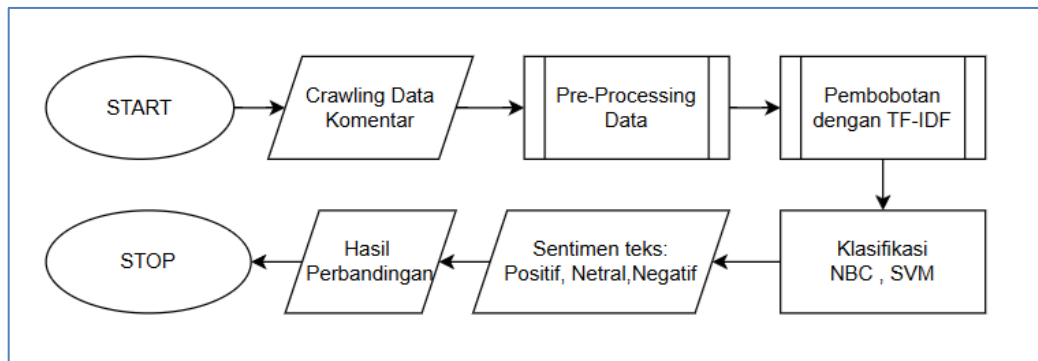
Instagram merupakan aplikasi media sosial yang sangat diminati oleh banyak kalangan untuk mengekspresikan diri. Bahkan di Indonesia pengguna Instagram adalah terbanyak ke-4 di dunia (Cindy, 2023), karena rentang pengguna yang sangat banyak, mulai dari orang awam, artis, sampai pejabat. Instagram mempunyai kelebihan karena memiliki banyak fitur antara lain berbagi konten secara interaktif, seperti *share*, pemberian komentar dan tanggapan komentar, mengunggah foto sampai video (Kartini dkk,2022). Dari sisi segmen pengguna, Instagram lebih banyak digunakan di kalangan pemuda, yaitu 71% pengguna berusia dibawah 30 tahun, 48% bersusia antara 30 sampai 40 tahun dan 29% pengguna antara usia 50 sampai 60 tahun (Anderson and Auxier,2021).

Penelitian tentang kinerja pemerintah atau kinerja Presiden melalui sentimen analysis telah banyak dilakukan oleh peneliti pada beberapa tahun terakhir. Yang pertama dilakukan oleh (Mahardika dan Yuliarso,2018) yang meneliti kinerja Pemerintah Joko Widodo berdasarkan data tweet menggunakan metode *Naïve Bayesian Classification* (NBC). Hasil klasifikasi memiliki akurasi 97%. Namun meskipun akurasi cukup tinggi, tetapi data yang digunakan cukup kecil, yaitu hanya 400 data. Penelitian kinerja Presiden melalui media Instagram juga dilakukan oleh (Rohim dan Pratama, 2022) dengan menggunakan algoritma NBC dengan data sebanyak 1152. Hasil yang diperoleh adalah Accuracy 83%, Precision 81%, Recall 84. Penelitian terbaru yang membandingkan kinerja presiden sebelum dan sesudah pemilihan presiden 2024 dilakukan oleh (Nehe dkk, 2024). Metode yang digunakan adalah metode NBC. Hasil penelitian menunjukkan penurunan sentiment positif setelah pemilihan presiden menurun dari 41,17% menjadi 39,08%. Namun dari berbagai penelitian kinerja presiden melalui media Instagram yang menggunakan perbandingan metode NBC dan *Support Vector Machine* (SVM) belum pernah dilakukan.

Penelitian ini akan melakukan analisis sentiment melalui klasifikasi teks komentar pengguna Instagram Presiden menggunakan metode NBC dan SVM. Kedua metode ini dipilih karena keunggulan dibanding metode klasifikasi yang lain seperti k-NN (Shedriko, 2021) atau metode *Decision Tree* (Shafri dan Ramle, 2009) Tujuan penelitian ini adalah mengetahui sejauh mana respon masyarakat terhadap kinerja Presiden pada bidang ekonomi. Penelitian juga bertujuan membandingkan kinerja algoritma NBC dan algoritma SVM. Metode klasifikasi NBC yang dirancang untuk digunakan dengan asumsi bahwa tiap variabel penjelas saling bebas, yang menunjukkan bahwa kehadiran atau ketiadaan suatu kejadian tertentu pada suatu kelompok tidak dikaitkan dengan kejadian lainnya (Lay, 2024) . Sedangkan metode *Support Vector Machine* (SVM) adalah metode prediksi dengan menemukan *hyperplane* terbaik yang membagi ruang input menjadi 2 atau lebih lebih klas (Ichwan dkk, 2018). Hasil penelitian ini akan menjadi opini alternatif penilaian kinerja Presiden oleh mesyarakat yang selama ini dilakukan oleh lembaga survei.

2. METODE PENELITIAN

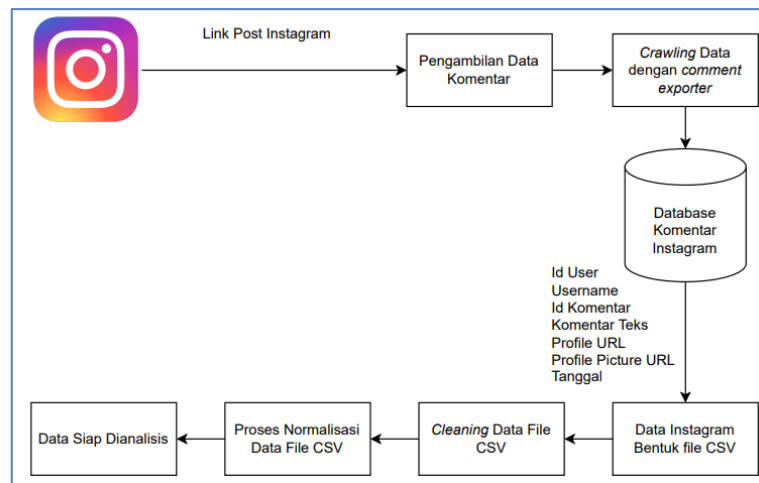
Metode penelitian ini menggunakan alur seperti pada Gambar 1 berikut ini.



Gambar 1. Diagram Alir Penelitian

2.1 Pengumpulan Data Komentar

Pengumpulan data diambil dari diambilkan dari data komentar pada akun Instagram Presiden <https://www.instagram.com/jokowi/>. Alur pengumpulan data adalah seperti pada Gambar 2 berikut ini.



Gambar 2. Proses *Crawling* data komentar

Adapun langkah-langkah pengumpulan data dan *pre-processing* data adalah sebagai berikut :

1. Dari Instagram diambil postingan yang relevan dengan kinerja presiden terutama bidang ekonomi dan industri. Dipilih tiga postingan yang relevan “Babak baru Industrialisasi Indonesia” (data-1), “Smelter PT Freeport Indonesia” (data-2) dan “Perkuat Ekonomi Syariah”(data-3)
2. Mengambil data komentar sebanyak 1563 untuk data-1, 2469 untuk data-2 dan 1032 untuk data-3.
3. Melakukan *crawling* menggunakan web *comment exporter* dan disimpan dalam bentuk csv.
4. Data csv memiliki struktur : *id user*, *username*, id komentar, komentar teks, *profile url*, *profile picture*, dan tanggal.
5. Proses *cleaning*, yaitu data yang diambil hanya data komentar saja, sehingga harus dilakukan pembersihan data dengan membuang *id*, *username* dan atribut lain yang tidak digunakan
6. Normalisasi data dengan cara merubah kata-kata yang tidak baku dan mengubah kata salah ketik menjadi benar.
7. Data siap dilakukan analisis dengan melakukan tahap *preprocessing* data dan klasifikasi data.

2.2 Tahap Pre-processing data

Pre-processing memiliki dua tahap, yaitu pre-processing dan pembobotan data, dengan rincian sebagai berikut:

1. Tahap *pre-processing* bertujuan membersihkan data yang tidak penting dalam pemrosesan data selanjutnya. *Pre-processing* memiliki beberapa tahap seperti, merubah semua kalimat menjadi huruf kecil, penghapusan karakter, angka, simbol, dan lain-lain, proses tokenisasi, dan proses pembuangan kata-kata stop word (*stopword removal*) menggunakan library Sastrawi.
2. Pembobotan TF-IDF bertujuan mengkonversi data teks menjadi data numerik dan diberi bobot pada tiap

katanya. TF-IDF mengubah teks data komentar ke dalam vektor yang semula berisi frekuensi kemunculan kata (TF) dalam teks tersebut. Selanjutnya dilakukan perhitungan IDF untuk mengetahui seberapa sering teks muncul pada dokumen yang berbeda yang diakhiri dengan proses perkalian nilai TF dan IDF agar memperoleh nilai bobot kata pada pada tiap teks tersebut.

2.3 Tahap Klasifikasi

Tahap berikutnya adalah melakukan klasifikasi data dengan menggunakan dua algoritma yaitu Naïve Bayes (NBC) Classifier dan *Support Vector Machine* (SVM).

Metode NBC

Metode NBC adalah teknik klasifikasi teks yang dengan menggunakan teori probabilitas Bayes dengan memandang kemunculan kata tertentu dalam suatu teks dapat diperhitungkan probabilitasnya setelah sebelumnya sederetan teks dalam klas-klas ditentukan probabilitasnya dalam tahap training (Lay,2024). Setelah mendapatkan hasil probabilitas awal (*prior probability*) dari kemunculan kata tiap kelas, maka klasifikasi teks kepada klas mana akan ditentukan dengan forumula berikut ini.

$$P(F|C) = \frac{P(F|C).P(C)}{P(F)} \quad (1)$$

di mana:

$P(C)$ = Probabilitas dari kelas C (*prior probability*)

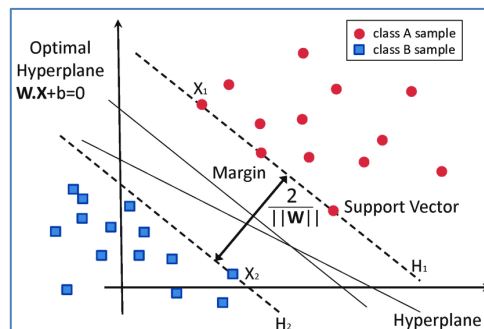
$P(F)$ = Probabilitas dari F (*evidence*)

$P(C|F)$ = Probabilitas kelas C berdasarkan kondisi F (*posterior probability*)

$P(F|C)$ = Probabilitas F berdasarkan kondisi kelas C

Metode *Support Vector Machine* (SVM)

Metode SVM merupakan metode dengan menganggap objek data yang diklasifikasi adalah kumpulan titik dalam ruang dimensi tinggi. Klasifikasi ditentukan dengan mencari suatu *hyperplane* yang dapat memisahkan kelompok objek dengan mencari jarak maksimal dari *hyperplane* (Gambar 3).



Gambar 3. *Hyperplane* dalam SVM

Klasifikasi ditentukan dengan memecahkan problem optimisasi pada persamaan *hyperplane* :

$$\text{Minimal } J_p = \frac{\|w\|^2}{2} + \sum_{i=1}^m \alpha_i [(w^T x_i + b)y_i - 1] \quad (2)$$

2.4. Metode Analisis Hasil Klasifikasi

Untuk melakukan analisis terhadap hasil kinerja algoritma klasifikasi, maka sebelum diklasifikasi perlu mengetahui sentimen dari komentar melalui tahap kegiatan pelabelan data. Teknik pemberian label dilakukan secara otomatis dengan menggunakan metode *lexicon-based* yang mengacu pada penelitian (Khaira dkk, 2020) . Teknik *lexicon-based* menggunakan library kata (kamus *lexicon*) yang telah diberikan label positif atau negatif dengan nilai +1 sampai +5 (untuk sentimen positif) dan nilai (-1 sampai -5) untuk sentimen negatif. Kata positif yang digunakan sebanyak 4405 kata dan kata negatif sebanyak 5209 kata. Berikut beberapa contoh kata pada kamus *lexicon* pada Tabel 1.

Tabel 1. Contoh Kamus Lexicon Kata Positif (kiri) dan Kata Negatif (kanan)

Kata	Skor sentimen	Kata	Skor sentimen
bagus	+1	rusak	-1
hebat	+5	bejat	-4
sempurna	+4	cemooh	-3
damai	+1	dendam	-2
gembira	+2	mampus	-5
senang	+3	mundur	-2

Untuk memberikan label pada komentar maka daftar token dari komentar yang telah melalui tahapan pre-processing dicek sentimennya di dalam kamus lexicon. Nilai sentimen dari tiap token kemudian dijumlahkan untuk mendapatkan nilai skor sentimen dari dokumen komentar. Jumlah skor sentimen dari setiap token digunakan untuk menetapkan nilai sentimen dari komentar dengan menggunakan aturanj, jika kalimat memiliki total skor < 0 maka kalimat memiliki label negatif, jika kalimat yang memiliki total skor > 0 maka kalimat memiliki label positif, dan jika kalimat yang memiliki total skor = 0 maka kalimat memiliki label netral.

Selanjutnya setelah proses klasifikasi data yang telah diberikan label dilakukan, pengujian apakah hasil klasifikasi sesuai dengan label atau tidak disusun matrik konfusi (confusion matrix) seperti pada Tabel 2. Matrik ini akan jadi ukuran bagi akurat tidaknya algoritma klasifikasi yang dilakukn dengan menghitung *recall*, *precision*, *accuration* dan *F1-score*.

Tabel 2. Tabel *Confusion Matrix*

		Nilai Aktual		
		Positif	Negatif	Netral
Prediksi	Positif	True Positive (TP)	False Positive (FP)	False Positif (FP)
	Negatif	False Negative (FN)	True Negative (TN)	False Negative (FN)
	Netral	False Neutral (FNeu)	False Neutral (FNeu)	True Neutral (TNeu)

Keterangan sel-sel matrik :

True Positive (TP) adalah data positif dan diprediksi benar sebagai positif.

False Positive (FP) adalah data negatif atau netral tetapi diprediksi sebagai positif.

False Negative (FN) adalah data positif atau netral tetapi diprediksi sebagai negatif.

True Negative (TN) adalah data negatif dan diprediksi benar sebagai negatif.

True Neutral (TNeu) adalah data yang diprediksi netral dan benar.

False Neutral (FNeu) adalah data yang diprediksi netral tetapi sebenarnya positif atau negatif

Berikut adalah persamaan parameter hasil klasifikasi yang diturunkan dari nilai *confusion matrix*, yaitu *accuracy* (*total accuracy*), *precision*, dan *recall* dengan rumus seperti persamaan

$$TA = \frac{TP + TNeu + TN}{TP + TNeu + TN + FP + FNeu + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

Dalam perhitungan *Precision* dan *Recall*, dilakukan rata-rata *precision* dan *recall* untuk setiap kelompok sentimen.

Untuk F1-Score adalah metrik evaluasi yang mengukur precision secara bersamaan :

$$F1 - score = \frac{2(Recall)(precision)}{Recall + Precision} \quad (6)$$

3. HASIL DAN PEMBAHASAN

3.1 Hasil ekstraksi data komentar

Instagram Presiden berisi berbagai postingan dengan keragaman topik yang sangat luas, mulai dari kegiatan presiden ketika liburan, kunjungan ke daerah, kunjungan ke luar negeri dan lain-lain. Untuk itu tidak semua diambil sebagai sumber data dan hanya kegiatan terkait perekonomian dan industri yang dijadikan sebagai sumber data penelitian ini. Untuk itu, maka data yang digunakan sebagai dataset untuk analisis dalam penelitian ini diambil dari 5 postingan yang terkait dengan bidang industri dan ekonomi yang memiliki komentar di atas 1000 komentar. Tabel 3 berikut ini judul posting dan jumlah komentar terkait tiap posting.

Tabel 3. Sumber Data Komentar dari Instagram Presiden

Tanggal Post	Judul Posting	Banyak Komentar
26-03-2024	Freeport Milik Indonesia	8.211
25-07-2024	Menuju IKN 2024	5.385
22-09-2024	Perkuat Eko Sistem Ekonomi Syariah	1.032
24-09-2024	Smelter PT Freeport Indonesia	2.468
29-09-2024	Babak baru Industrialisasi Indonesia	1.564
Jumlah total komentar		18.660

Setelah memilih postingan, pengambilan data komentar dengan cara menyalin *link* postingan dan menggunakan ekstensi *comment exporter* yang telah di-*install* untuk mengubah data komentar tersebut menjadi file csv. Sebelum dilakukan *pre-processing*, langkah pertama adalah membuang komentar yang tidak berisi kata-kata dan hanya berisi emoticon seperti : 🤔🔥🔥🔥, 🤔🤔🤔🤔🤔🤔🤔🤔🤔, 🤔🤔🤔🤔🤔🤔🤔🤔🤔. Meskipun *emoticon* tersebut memuat sentimen positif atau negatif, tetapi komentar yang hanya berisi *emoticon* terpaksa dibuang karena di luar jangkauan penelitian ini, sehingga data yang diambil hanya murni data komentar. Setelah dipilih hanya data komentar yang digunakan, maka diperoleh koleksi data komentar sebanyak 15.306 komentar.

Selanjutnya seluruh data mentah yang diperoleh kemudian diolah dengan dilakukan langkah-langkah pada tahap *preprocessing*. Contoh hasil langkah-langkah *pre-processing* suatu komentar original :

“Tabung Gas LPG langka pak. 🤔 sama jalan di pedesaan amburadul!!! 10 juta lapangan kerja baru mana?”

Data Cleaning

Pada tahap ini, teks akan di-identifikasi dan dibersihkan dari atribut kata yang tidak diperlukan serta kata duplikat. Tabel 4 berikut memperlihatkan perbedaan sebelum dan sesudah tahapan-tahapan *pre-processing*.

Tabel 4. Hasil proses data cleaning

Teks original	Data cleaning
Tabung Gas LPG langka pak. 🤔 sama jalan di pedesaan amburadul!!! 10 juta lapangan kerja baru mana?	Tabung Gas LPG langka Pak sama jalan di pedesaan amburadul juta kerja baru mana

Case folding

Tahapan selanjutnya, data diubah menjadi huruf (*lowercase*). Tabel 5 berikut ini adalah tabel untuk memperlihatkan perbedaan sebelum dan sesudah *case folding* diubah.

Tabel 5. Hasil proses *case folding*

<i>Data cleaning</i>	<i>Case folding</i>
Tabung Gas LPG langka pak sama jalan di pedesaan amburadul juta lapangan kerja baru mana	tabung gas lpg langka pak sama jalan di pedesaan amburadul juta kerja baru mana

Stop Word Filtering

Tahap selanjutnya, semua data yang terkategori stop word ('ini', 'yang', 'dan', dst) difilter. Tabel 6 berikut adalah tabel untuk memperlihatkan perbedaan dari case folding ke Stop word filtering.

Tabel 6. Hasil proses *stop word filtering*

<i>Case Folding</i>	<i>Stop Word Filtering</i>
tabung gas lpg langka pak sama jalan di pedesaan amburadul juta kerja baru mana	tabung gas lpg langka pak sama jalan pedesaan amburadul juta kerja baru mana

Stemming

Tahap berikutnya adalah proses stemming, yaitu mengembalikan kata ke akar kata menggunakan stemmer. Tabel 7 berikut adalah tabel untuk memperlihatkan perbedaan setelah stemming.

Tabel 7. Hasil proses *stemming*

<i>Stop Word Filtering</i>	<i>Stemming</i>
tabung gas lpg langka pak sama jalan pedesaan amburadul juta kerja baru mana	tabung gas lpg langka pak sama jalan desa amburadul juta kerja baru mana

Tokenization

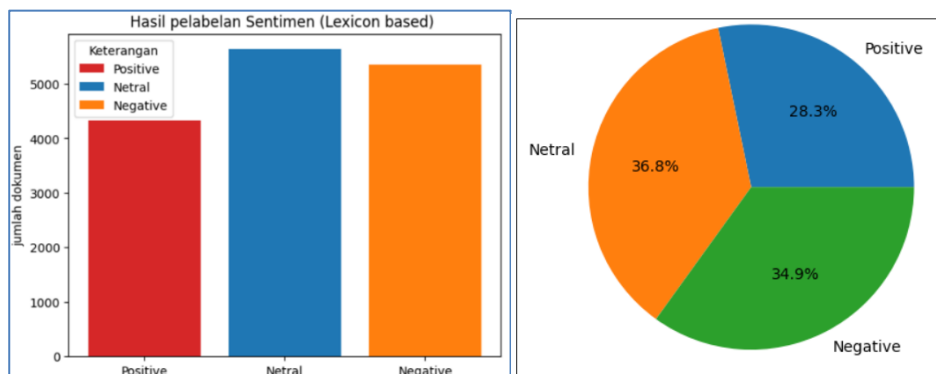
Tahap terakhir dari pre-processing adalah proses tokenization, yakni merubah teks menjadi deretan token yang akan dihitung frekuensi kemunculannya pada tahap analisis TF-IDF. Tabel 8 berikut adalah hasil *tokenization*.

Tabel 8. Hasil proses *tokenization*

<i>Stemming</i>	<i>Tokenization</i>
tabung gas lpg langka pak sama jalan desa amburadul juta kerja baru mana	'tabung', 'gas', 'lpg', 'langka', 'sama', 'jalan' 'desa' , 'amburadul', 'juta', 'kerja', 'baru' mana

3.2 Hasil klasifikasi data komentar

Data yang telah berhasil dilabeli menggunakan *lexicon based* adalah sebanyak 15.306 komentar. Hasil pelabelan disajikan dalam Gambar 4 berikut ini. Dari gambar tersebut terlihat hasil pelabelan adalah 28.3% positive, 36.8% netral dan 34.9% negatif.



Gambar 4. Hasil pelabelan data dengan *Lexicon Based*

Berdasarkan analisis penggunaan kata yang sering muncul diambil 30 kata dengan frekuensi kemunculan tertinggi dalam setiap kelas sentimen, dapat disajikan grafik *word count* seperti pada Gambar 5 berikut, tanpa menampilkan *word count* untuk opini netral.



Gambar 5. *Word Count* untuk opini positif (kiri) dan opini negatif (kanan)

Hasil klasifikasi dengan menggunakan algoritma NBC dan algoritma SVM disajikan analisisnya pada Tabel 9 dan Tabel 10.

Tabel 9. Hasil Klasifikasi Metode NBC

No.	Rasio Training:Testing	Accuracy	Precision	Recall	F1-score
1.	80:20	67.1%	68.2%	66.9%	63.5%
2.	70:30	68.7%	68.5%	67.8%	68.2%
3.	60:40	64.6%	64.7%	65.0%	62.4%
4.	50:50	65.4%	66.2%	65.2%	65.8%

Tabel 10. Hasil Klasifikasi SVM

No.	Rasio Training:Testing	Accuracy	Precision	Recall	F1-score
1.	80:20	79.1%	78.2%	78.3%	78.3%
2.	70:30	80.5%	80.5%	79.6%	79.6%
3.	60:40	78.2%	78.6%	78.2%	78.4%
4.	50:50	77.4%	77.2%	76.7%	77.6%

3.3 Hasil Perbandingan Klasifikasi NBC dan SVM

Setelah masing-masing algoritma dilakukan training dan testing pada 4 ombinasi training:testing , yaitu 80:20, 70:03,60:40 dan 50:50 dan dihitung parameter-parameter klasifikasi berdasarkan confusion matrix yang terbentuk, maka dicari komposisi yang terbaik. Tabel 11 berikut adalah hasil perbandingan dari kinerja dari metode NBC dan SVM pada komposisi yang paling baik 70:30

Tabel 2 Perbandingan Kinerja algoritma NBC dan SVM

	Akurasi	Precision	Recall	F1-score
Klasifikasi NBC	68.7%	68.5%	67.8%	68.2%
Klasifikasi SVM	80.5%	80.5%	79.6%	79.6%

Dari tabel tersebut terlihat bahwa pada seluruh parameter klasifikasi berupa accuracy, recall dan precision serta F1-score, algoritma SVM mendapatkan kinerja yang lebih baik dari pada algoritma NBC.

4. KESIMPULAN

Dari dataset komentar masyarakat pada 5 program bidang ekonomi dan industri didapatkan sebanyak 18.660 komentar. Setelah dilakukan filtering dengan membuang komentar yang hanya berupa emoticon diperoleh sebanyak 15.306 komentar. Dari dataset berupa 15.306 komentar dilakukan pelabelan menggunakan metode lexicon based, yaitu menentukan sentimen komentar berdasarkan sentimen dari kata-kata penyusunnya, diperoleh

sentimen komentar yaitu 28.3% positive, 36.8% netral dan 34.9% negatif. Hasil klasifikasi menggunakan dua metode NBC dan SVM didapatkan parameter klasifikasi berupa *accuracy* sebesar 68.7%, *precision* sebesar 68.5%, *recall* sebesar 67.8% dan *F1-score* 68.2% untuk algoritma NBC. Untuk algoritma SVM menghasilkan *accuracy* sebesar 80.5%, nilai *precision* 80.5%, nilai *recall* 79.6%, dan nilai *F1-score* 79.6%. Dapat disimpulkan bahwa dari dataset yang ada terlihat bahwa algoritma SVM lebih unggul dari pada algoritma NBC pada semua parameter klasifikasi. Dari beberapa komposisi training dan testing juga didapatkan komposisi terbaik adalah komposisi 70:30.

UCAPAN TERIMA KASIH

Pada kesempatan ini penulis mengucapkan terimakasih kepada semua pihak yang telah terlibat dalam penyusunan naskah dan proses penelitian ini.

DAFTAR PUSTAKA

- Anderson, M. and Auxier, B. (2021). "Social Media Use in 2021," *Pew Res. Cent.*, no. April, pp. 1–6, 2021, [Online]. Available: <https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/>
- Chaterine, R. H. dan Prabowo, D. (2024). <https://nasional.kompas.com/read/2024/10/04/16081911/survei-indikator-mayoritas-masyarakat-puas-kinerja-jokowi-selama-2-periode>. di ases 5 Oktober 2024.
- Cindy, A. M. (2023). "Jumlah Pengguna Instagram Indonesia Terbanyak ke-4 di Dunia," *Databoks.Katadata*, vol. 2022, p. 1, [Online]. Available: <https://databoks.katadata.co.id/datapublish/2023/05/04/jumlah-pengguna-instagram-indonesia-terbanyak-ke-4-di-dunia>
- Fadiyah, D. dan Simorangkir, S. (2019). Penggunaan Media Sosial Instagram dalam Membangun Citra Positif Presiden Joko Widodo pada Pilpres 2019. *Jurnal of Political Issues*. Vol.3. No.1. pp.13-27
- Ichwan, M., Dewi, I. A. and Muharom, Z. (2018). Klasifikasi Support Vector Machine (SVM) Untuk Menentukan Tingkat Kemanisan Mangga Berdasarkan Fitur Warna. *MIND Journal*. Vol.3.No.2, pp. 16-24
- Kartini, J. Syahrina, N. Siregar, and N. Harahap, "Penelitian Tentang Instagram," *Maktab. J. Perpust. dan Inf.*, vol. 2275, pp. 20–26, 2022, [Online]. Available: <https://ummaspul.e-journal.id/RMH/article/view/4466>
- Khaira, U. R., Johanda, P. E. P. Utomo, and T. Suratno. (2020). "Sentiment Analysis Of Cyberbullying On Twitter Using SentiStrength," *Indones. J. Artif. Intell. Data Min.*, vol. 3, no. 1, p. 21-29
- Lay, R. (2024). Naive Bayes Classifier Explained: Applications and Practical Problems. Diases tanggal 20 Oktober 2024 dari <https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/>
- Mahardika, Y. S. dan Zuliarso, E. (2018). Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial Twitter Menggunakan Algoritma Naives Bayes Classifier. *Prosiding SINTAK*. Pp-417-429
- Mandala, A. (2024). Peranan Media Sosial dalam Membangun Citra Positif Lembaga Pemerintah. *Prosiding Seminar Nasional Komunikasi (SENAKOM)*. Vol.1.No.1
- Mao, Y., Liu, Q. and Zhang, Y. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. *Journal of King Saud University - Computer and Information Sciences*, Vol.36.issue.4,231-245
- Mukhroman, I., Ahmad, I. dan Gumelar R. G. (2024). Pembentukan Opini Publik Pada Pemilihan Umum 2024 (Studi Kasus Opini Publik – Survey Elektabilitas Capres-Cawapres 2024 Pasca Debat Pertama). *JOURNAL SYNTAX IDEA*. Vol 6 No.6. 2472-2482.
- Nehe, P. H., Berutu, S. S. dan Budiati, H. (2024). Analisis Sentimen Opini Masyarakat Terhadap Presiden Jokowi Sebelum Dan Sesudah Pilpres 2024 Menggunakan Metode Naive Bayes Classification. *JUTISI: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*. Vol 13.No.1. pp 451-463
- Nufus, W. H. (2024). <https://news.detik.com/berita/d-7588975/survei-lsi-denny-ja-80-8-responden-puas-kinerja-presiden-jokowi>. Diases 16 Oktober 2024
- Rohim, A. N. dan Pratama, A. R. (2022) Analisis Sentimen Publik di Media Sosial Instagram atas Kinerja Presiden Joko Widodo. *Automata*. Vol 3. No.2.
- Shafri, H. Z. M. dan Ramle, F. S. H. (2009). Perbandingan Klasifikasi Support Vector Machine dan Decision Tree Menggunakan Data Satelit Pulau Langkawi. *Information Technology Journal*, 8: 64-70.
- Shedriko (2021). Perbandingan Algoritma SVM dan k-NN dalam Mengklasifikasi Kelulusan Mahasiswa Pada Suatu Mata Kuliah. *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*. Vol. 6 No. 2 D115-128.