



Analisis Sentimen Pada Media Sosial *Instagram* Terhadap Akun Presiden Joko Widodo Menggunakan Metode *Naïve Bayes Classifier*

Della Berliansyah

Universitas Muhammadiyah Jember

Ulya Anisatur

Universitas Muhammadiyah Jember

Habibatul Azizah Alfaruq

Universitas Muhammadiyah Jember

Alamat: Jl. Karimata No.49 Jember

Korespondensi penulis: dellaberlian8@gmail.com

Abstract

In the growing digital era, social media, especially Instagram, has become the main platform for people to communicate and express themselves. One of the most influential accounts is President Joko Widodo's official account, @jokowi, which is often in the spotlight with thousands of comments covering a wide range of sentiments, both positive and negative. In the midst of his popularity, sentiment analysis is key to understanding the public's views on Jokowi's leadership. This study aims to analyze public sentiment towards President Joko Widodo (Jokowi) through comments posted on his official Instagram account (@jokowi). By utilizing the Naïve Bayes Classifier method, this study collected data from 1000 comments which were then processed through various stages of the methodology, including data collection, preprocessing, weighting, k-fold cross validation, and method implementation. Through the preprocessing stage involving cleansing, stopword removal, stemming, and tokenizing, the comments were prepared for further analysis. Test results using k-fold cross validation show that the model has an average accuracy of 80.3%. In addition, evaluation using confusion matrix showed an accuracy of 84.1%, with a precision of 85.5% and recall of 92.4%. These results show that the Naïve Bayes Classifier method performs well in classifying positive and negative sentiments in the comments.

Keywords: Accuracy, Instagram, Jokowi, Naïve Bayes Classifier, Sentiment Analysis.

Abstrak

Dalam era digital yang kian berkembang, media sosial, khususnya Instagram, telah menjadi platform utama bagi masyarakat untuk berkomunikasi dan berekspresi. Salah satu akun yang sangat berpengaruh adalah akun resmi Presiden Joko Widodo, @jokowi, yang sering menjadi sorotan dengan ribuan komentar yang mencakup beragam sentimen, baik positif maupun negatif. Di tengah popularitasnya, analisis sentimen menjadi kunci untuk memahami pandangan masyarakat terhadap kepemimpinan Jokowi. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap Presiden Joko Widodo (Jokowi) melalui komentar-komentar yang diposting di akun resmi Instagramnya (@jokowi). Dengan memanfaatkan metode *Naïve Bayes Classifier*, penelitian ini mengumpulkan data dari 1000 komentar yang kemudian diproses melalui berbagai tahapan metodologi, termasuk pengumpulan data, preprocessing, pembobotan, k-fold cross validation, dan implementasi metode. Melalui tahap preprocessing yang melibatkan cleansing, stopword removal, stemming, dan tokenizing, komentar-komentar dipersiapkan untuk analisis lebih lanjut. Hasil pengujian menggunakan k-fold cross validation menunjukkan bahwa model memiliki rata-rata akurasi sebesar 80,3%. Selain itu, evaluasi menggunakan *confusion matrix* menunjukkan akurasi sebesar 84,1%, dengan presisi sebesar 85,5% dan recall sebesar 92,4%. Hasil ini menunjukkan bahwa metode *Naïve Bayes Classifier* memiliki kinerja yang baik dalam mengklasifikasikan sentimen positif dan negatif dalam komentar-komentar tersebut.

Kata Kunci: Akurasi, Analisis Sentimen, Instagram, Jokowi, *Naïve Bayes Classifier*.

LATAR BELAKANG

Media sosial merupakan *platform* digital yang banyak digunakan untuk berkomunikasi dalam bentuk teks, foto, audio, dan video. Contoh dari media sosial yaitu *facebook*, *twitter*, *instagram*, *whatsapp*, dan lain sebagainya. Pengguna media sosial terbanyak kedua di Indonesia adalah *Instagram*. Penggunaannya meliputi berbagai kalangan, mulai dari anak-anak hingga orang dewasa, yang berkontribusi pada tingginya popularitas *instagram*. Berdasarkan Napoleon Cat 89,64 juta pengguna *instagram* pada Desember 2023. Dalam setiap unggahan *instagram*, setiap individu dapat dengan bebas mengungkapkan pendapatnya dengan menulis komentar terkait unggahan orang lain (Maulidina, 2020).

Penggunaan Instagram oleh lembaga pemerintahan, termasuk akun resmi Presiden Indonesia Joko Widodo (@jokowi), menjadi peluang untuk memberikan layanan, kebijakan publik, dan mengumpulkan masukan dari masyarakat. Akun ini sering menjadi tempat beragam komentar, mencakup kritik, pujian, dan hinaan, memberikan gambaran opini masyarakat terhadap kinerja presiden. Unggahan, seperti yang membahas sambutan hangat dari masyarakat Kutai Barat, dapat menerima ribuan komentar, mencerminkan beragam pandangan terhadap kepemimpinan Jokowi. Sebagai presiden ketujuh Indonesia, Jokowi dianggap sebagai tokoh modern yang membawa model politik baru, mendekati diri kepada anak muda, dan menganut nilai-nilai Pancasila sambil berpihak pada rakyat.

Fitur komentar memungkinkan pengguna untuk mengekspresikan opini mereka terhadap gambar atau video yang diunggah, baik dalam bentuk kalimat terstruktur maupun tidak terstruktur. Meskipun pengguna memiliki hak untuk berkomentar, perlu dilakukan analisis sentimen untuk memproses teks dan mendapatkan informasi yang akurat. Analisis ini menjadi kunci untuk mengklasifikasikan dan mengategorikan komentar, membantu dalam menilai pandangan masyarakat secara lebih baik.

Analisis sentimen melibatkan proses mengidentifikasi dan mengelompokkan pendapat dalam bentuk teks ke dalam kategori sentimen positif atau negatif. Pentingnya analisis sentimen terletak pada kemampuannya untuk menganalisis beragam aspek yang mencakup pendapat dan perasaan masyarakat terhadap topik tertentu. Sumber informasi utama untuk analisis sentimen berasal dari media sosial, yang merupakan *platform* yang digunakan secara luas oleh masyarakat (Rusdaman & Rosiyadi, 2019).

Salah satu teknik pembelajaran dari *text mining* untuk melakukan analisis sentimen adalah *Naïve Bayes Classifier*. Metode *Naïve Bayes Classifier* juga dapat digunakan pada domain yang berbeda. Metode ini dianggap memiliki potensi yang baik untuk melakukan

klasifikasi data dibandingkan dengan metode klasifikasi lainnya, baik dalam hal akurasi maupun komputasi (Watraton et al., 2020).

Penelitian ini bertujuan untuk memanfaatkan platform Instagram dengan menganalisis komentar berbahasa Indonesia yang membahas Presiden Jokowi pada akun resminya. Fokus penelitian adalah menetapkan persentase tanggapan dan komentar dari beberapa akun terhadap unggahan @jokowi, yang dapat menjadi masukan untuk meningkatkan kualitas kinerja pemerintahan atau sebagai evaluasi. Dengan melakukan analisis sentimen menggunakan metode Naïve Bayes Classifier, penelitian ini bertujuan memahami pendapat publik di Instagram terhadap Presiden Jokowi, baik positif maupun negatif. Tugas akhir ini berjudul "Analisis Sentimen Pada Media Sosial Instagram Terhadap Akun Presiden Joko Widodo Menggunakan Metode Naïve Bayes Classifier" sebagai respons terhadap permasalahan yang dijelaskan sebelumnya.

KAJIAN TEORITIS

A. Presiden Joko Widodo

Menurut (Ardiansyah, 2021) Ir. H. Joko Widodo, yang dilahirkan di Surakarta, Jawa Tengah, pada 21 Juni 1961, merupakan Presiden ke-7 Republik Indonesia. Memulai masa jabatannya pada 20 Oktober 2014. Sebelumnya, Joko Widodo terlibat dalam pemerintahan sebagai Wali Kota Surakarta (Solo) dari 28 Juli 2005 hingga 1 Oktober 2012.

B. Instagram

Instagram adalah *platform* media sosial yang dimiliki oleh *Meta Platforms*, perusahaan Amerika. Layanan ini memungkinkan pengguna untuk berbagi foto dan video, yang dapat diedit dengan berbagai *filter*, serta diberi tagar dan lokasi. Pada awalnya, *Instagram* memiliki pembatasan konten dengan rasio aspek persegi (1:1) dan lebar 640. Namun, pada tahun 2015, batasan ini diperluas menjadi 1080 piksel (Nisa, 2023).

C. Sentimen Analisis

Sentimen analisis, merupakan penambangan opini, melibatkan identifikasi sikap terhadap entitas atau individu tertentu. Dalam analisis yang lebih mendalam, kita dapat mengidentifikasi sumber-sumber sentimen positif dan negatif, baik individu maupun kelompok yang terlibat (Widayat, 2021).

D. Python

Python, yang dibuat oleh Guido van Rossum di Belanda pada tahun 1990, telah menjadi bahasa pemrograman yang sangat populer digunakan oleh perusahaan besar dan pengembang

untuk menciptakan berbagai aplikasi desktop, web, dan mobile. Nama "Python" sendiri diambil dari acara televisi favorit Guido, yaitu *Monty Python's Flying Circus*.

E. Google Colab

Dijelaskan oleh Noviah, (2023) Google Colaboratory, atau yang dikenal sebagai Google Colab, adalah dokumen yang dapat dieksekusi yang memungkinkan pengguna untuk menyimpan, menulis, serta berbagi program yang telah mereka buat melalui Google Drive. Dengan Google Colab, pengguna dapat dengan mudah mengimpor dan menginstal berbagai pustaka Python terkenal, seperti TensorFlow, PyTorch, Pandas, dan lainnya (Thoyyibah et al., 2024).

F. Text mining

Text mining (atau juga disebut *text analytics*) adalah proses ekstraksi informasi berharga atau pengetahuan yang terkandung dalam teks atau dokumen teks. Dijelaskan juga oleh (Firdaus & Firdaus, 2021) tentang *Text mining* adalah proses *ekstraksi* informasi dari teks yang berasal dari dokumen tertentu.

G. Text Preprocessing

Text preprocessing adalah serangkaian langkah atau teknik yang digunakan untuk membersihkan, mempersiapkan, dan mengolah teks sebelum dilakukan analisis lebih lanjut atau pemrosesan bahasa alami (*Natural Language Processing*, NLP) (Isnain et al., 2021).

H. Algoritma TF-IDF

Menurut Sari et al., 2021 pembobotan *TF/IDF* adalah sebuah teknik yang memberikan nilai bobot kepada kata-kata dalam suatu dokumen dengan memperhitungkan dua konsep utama: frekuensi kemunculan kata dalam dokumen tertentu disebut *Term Frequenc (TF)* dan frekuensi keberadaan kata dalam seluruh koleksi dokumen disebut *Inverse Document Frequency* atau *IDF*.

I. K-Fold Cross Validation

Cross-Validation dikenal sebagai estimasi rotasi adalah metode validasi model. Misalnya, pada percobaan pertama, subset S1 berperan sebagai data uji, sementara subset yang lain menjadi data pelatihan. Pada percobaan kedua, subset S1, S3, ..., Sk menjadi data pelatihan, sedangkan S2 menjadi data uji, dan seterusnya (Tempola et al., 2018).

Metode *cross validation* digunakan untuk mengevaluasi keberhasilan suatu model dengan menerapkan perulangan dengan komposisi masing-masing dari data latih dan data uji sehingga algoritma yang digunakan diuji validitasnya. Namun, validasi silang *k-fold* adalah metode validasi silang yang menggunakan pembagian data ke dalam subset data *k* dengan

pembagian jumlah yang seimbang. Selama proses ini, *testing* dan *training* dilakukan berulang kali. Data dibagi menjadi k komponen (Zhafira et al., 2021).

J. *Naïve Bayes Classifier*

Naïve Bayes Classifier adalah metode *Classifier* yang berdasarkan probabilitas dan teorema Bayesian dengan asumsi keindependenan atribut. Algoritma *Naïve Bayes Classifier* terdiri dari dua tahap yaitu tahap pertama melibatkan pelatihan menggunakan himpunan dokumen contoh atau data latih, sementara tahap kedua melibatkan proses klasifikasi dokumen (Widowati & Sadikin, 2021).

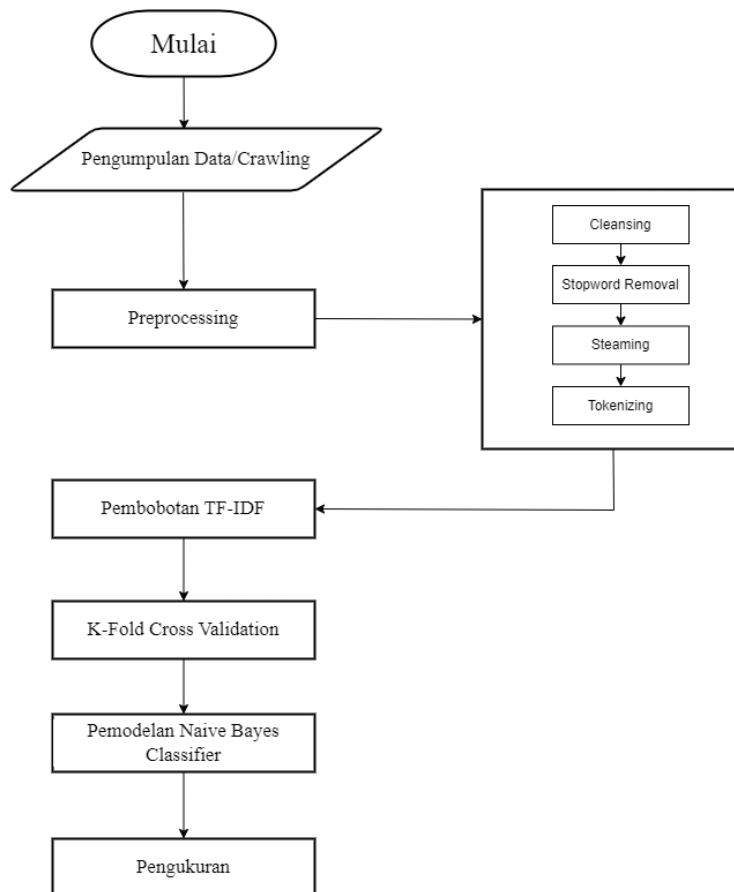
K. *Confusion Matrix*

Confusion Matrix adalah alat yang berguna dalam mengevaluasi model klasifikasi dengan menghitung jumlah data uji yang diklasifikasikan dengan benar dan yang salah (Normawati & Prayogi, 2021). Menurut Azhari et al., 2021, *True positive*, *false positive*, *false negatif* dan *true negative*.

METODE PENELITIAN

Penelitian tentang analisis sentimen pada Instagram terhadap Presiden Joko Widodo dengan Metode *Naïve Bayes Classifier* melibatkan beberapa tahapan metodologi penelitian yang sistematis. Tahapan tersebut meliputi pengumpulan data, preprocessing, pembobotan, k-fold cross validation, implementasi metode *Naïve Bayes Classifier*, dan pengukuran hasil. Pengumpulan data dilakukan melalui crawling komentar pada postingan Instagram Presiden Joko Widodo. Kemudian, data tersebut melalui tahap preprocessing seperti cleansing, stopword removal, stemming, dan tokenizing. Selanjutnya, dilakukan pembobotan menggunakan metode TF-IDF. Proses k-fold cross validation digunakan untuk evaluasi hasil dengan membagi data menjadi subset untuk training dan testing. Implementasi metode *Naïve Bayes Classifier* melibatkan pembagian kata ke dalam kategori sentimen positif dan negatif, diikuti oleh perhitungan probabilitas dan klasifikasi. Pengukuran hasil dilakukan menggunakan confusion matrix untuk menentukan akurasi, presisi, dan recall. Hasil penelitian menunjukkan tingkat akurasi sebesar 70%, presisi 80%, dan recall 66.67%.

Adapun tahapan penelitian yang digunakan untuk analisis sentimen pada *instagram* terhadap presiden Joko Widodo dengan Metode *Naïve Bayes Classifier* memiliki langkah-langkah yang dapat digambarkan seperti berikut:



Gambar 3. Tahapan Penelitian

1. **Mulai:** Penelitian dimulai dengan pengumpulan data dari komentar-komentar yang terdapat pada akun Instagram resmi Presiden Joko Widodo (@jokowi).
2. **Pengumpulan data/crawling:** Data komentar diambil melalui proses crawling pada postingan akun @jokowi, untuk digunakan dalam analisis sentimen.
3. **Preprocessing:** Data yang telah terkumpul kemudian melewati tahap preprocessing, yang mencakup beberapa proses seperti cleansing, stopword removal, stemming, dan tokenizing.
4. **Dalam preprocessing (cleansing, stopword removal, stemming, tokenizing):**
 - a. **Cleansing:** Proses menghilangkan karakter atau informasi yang tidak relevan atau berpengaruh terhadap analisis sentimen.
 - b. **Stopword removal:** Langkah untuk menghapus kata-kata umum yang tidak memberikan kontribusi signifikan terhadap analisis sentimen.
 - c. **Stemming:** Proses mengubah kata-kata menjadi bentuk dasarnya atau menghapus imbuhan sehingga mempermudah analisis.

- d. **Tokenizing:** Memisahkan kata-kata dalam setiap komentar untuk mempersiapkan data agar dapat diolah lebih lanjut.
5. **Pembobotan TF-IDF:** Setelah preprocessing, dilakukan pembobotan kata menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) untuk memberi bobot pada setiap kata dalam setiap komentar.
6. **K-Fold Cross Validation:** Data kemudian dibagi menjadi beberapa subset menggunakan metode K-Fold Cross Validation untuk mengevaluasi hasil analisis dengan membagi data menjadi subset untuk pelatihan (training) dan pengujian (testing).
7. **Pemodelan Naive Bayes Classifier:** Implementasi metode Naive Bayes Classifier dilakukan untuk mengklasifikasikan komentar-komentar berdasarkan sentimen positif atau negatif.
8. **Pengukuran:** Terakhir, dilakukan pengukuran hasil analisis menggunakan metrik-metrik seperti akurasi, presisi, dan recall untuk mengevaluasi performa model Naive Bayes Classifier dalam memprediksi sentimen dari komentar-komentar tersebut.

HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Data yang digunakan pada penelitian ini merupakan komentar dari pengguna instagram yang ada pada postingan akun @jokowi. Data yang dikumpulkan menggunakan teknik *crawling* dengan *plugin web browser google chrome scraper (IGCommentExport)* dari awal terbuatnya akun tersebut hingga saat ini dengan total *crawling* 100.000 komentar dan yang digunakan di penelitian ini sebanyak 1000 komentar.

User Id	Username	Comment Id	Comment Text	Profile URL	Avatar URL	Date
1,71E+09	ranie.hassa	1,80499E+16	@joentiq woow bicaranya , Pak Jokowi bayar ora	https://www.instagram.com,	https://instagram.fsub8-1.f11/14/2023,	5:37:49 AM
1,71E+09	ranie.hassa	1,7917E+16	@indahwahyuni.iw80 🍀🍀🍀🍀	https://www.instagram.com,	https://instagram.fsub8-1.f11/14/2023,	5:40:21 AM
4,7E+09	indraarieirz	1,7920E+16	@ranie.hassan yakin akun2 hoax bicara mu kek s	https://www.instagram.com,	https://instagram.fsub8-2.f11/14/2023,	5:44:40 AM
1,71E+09	ranie.hassa	1,80255E+16	@indahwahyuni.iw80 🍀🍀	https://www.instagram.com,	https://instagram.fsub8-1.f11/14/2023,	5:52:22 AM
1,71E+09	ranie.hassa	1,79978E+16	@zefano_shop kata siapa Pak Jokowi gak pernah	https://www.instagram.com,	https://instagram.fsub8-1.f11/14/2023,	5:57:12 AM
1,71E+09	ranie.hassa	1,80126E+16	@zefano_shop bukankah sekrgn Buzzer2 PDlp ju	https://www.instagram.com,	https://instagram.fsub8-1.f11/14/2023,	5:59:12 AM
1,71E+09	ranie.hassa	1,78463E+16	Emang keluanku bagaimana ?	https://www.instagram.com,	https://instagram.fsub8-1.f11/14/2023,	6:00:45 AM
4,7E+09	indraarieirz	1,80174E+16	@ranie.hassan keluan?? Belajar abjad dulu baru	https://www.instagram.com,	https://instagram.fsub8-2.f11/14/2023,	6:06:36 AM
1,71E+09	ranie.hassa	1,80019E+16	Halah salah dikit aja jadi riweuh banget 😂😂😂	https://www.instagram.com,	https://instagram.fsub8-1.f11/14/2023,	6:53:26 AM
4,7E+09	indraarieirz	1,79233E+16	@ranie.hassan mang eak	https://www.instagram.com,	https://instagram.fsub8-2.f11/14/2023,	7:06:07 AM
1,71E+09	ranie.hassa	1,80086E+16	Weewe aaah 😊	https://www.instagram.com,	https://instagram.fsub8-1.f11/14/2023,	9:15:56 AM
6,21E+10	4168johan	1,83836E+16	Indonesia luas pak bukan di kutai barat aja	https://www.instagram.com,	https://instagram.fblr20-3.11/14/2023,	9:21:48 AM
2,63E+08	erza_irlfan4	1,80032E+16	@joe.zebug loe bilang mata gwe minus??? Seena	https://www.instagram.com,	https://instagram.fsub32-1.11/14/2023,	11:11:56 AM
5,84E+09	faidwahyuc	1,79889E+16	@agus.n99 kami? kami gundulmu atos	https://www.instagram.com,	https://instagram.fsub8-1.f11/14/2023,	1:39:09 PM
6,14E+10	yustina279	1,79867E+16	🍀	https://www.instagram.com,	https://instagram.fsub32-1.11/15/2023,	7:43:59 PM
6,21E+10	amaratbha	1,83244E+16	🍀🍀🍀🍀🍀🍀	https://www.instagram.com,	https://instagram.fsub8-2.11/15/2023,	9:57:22 PM
5,3E+10	syifasyifa81	1,80248E+16	@joentiq Aku. Nggak ngurusin politik. Yang kura	https://www.instagram.com,	https://instagram.fsub32-2.11/16/2023,	4:38:45 AM
6,2E+10	xyzfakta	1,8212E+16	@ulfaazzam2 🍀🍀Jokowi & times Jokowi : Kita	https://www.instagram.com,	https://instagram.fsub8-1.f11/16/2023,	7:46:56 AM
5,03E+10	ulfaazzam2	1,81072E+16	@xyzfakta saya gada ikut bgtuan. Tim sukseslah	https://www.instagram.com,	https://instagram.fsub8-2.f11/16/2023,	10:07:49 AM
1,41E+09	isminuurist	1,82754E+16	@indahwahyuni.iw80 🍀🍀 bener juga lulusan sn	https://www.instagram.com,	https://instagram.fsub8-1.f11/17/2023,	4:12:37 AM
6,29E+10	monica76b	1,79893E+16	🍀	https://www.instagram.com,	https://instagram.fsub8-1.f11/17/2023,	11:26:12 AM
5,51E+10	rqgly_p	1,83094E+16	@agus.n99 kayak lagu cuman kamuuuuuuu huu	https://www.instagram.com,	https://instagram.fsub8-1.f11/17/2023,	3:13:17 PM
6,08E+10	crea7live.sti	1,82455E+16	Pak, lagi rame berita di Aceh banyak orang rohin	https://www.instagram.com,	https://instagram.fsub8-2.f11/17/2023,	9:03:17 PM
4,98E+09	syamsul.asl	1,78582E+16	@agus.n99 pretlah...	https://www.instagram.com,	https://instagram.fblr20-3.11/18/2023,	9:58:11 AM
3,11E+09	faryd_abd	1,80004E+16	@agus.n99 kami? Lu aja kali 🍀	https://www.instagram.com,	https://instagram.fsub32-2.11/18/2023,	5:42:26 PM
2,04E+09	zeindcc	1,79944E+16	@agus.n99 ngomong opo atuh guss.. aguss aguss	https://www.instagram.com,	https://instagram.fblr20-3.11/18/2023,	9:00:13 PM

Gambar 4. Hasil *Crawling* Data Komentar

Pada gambar 4.1 merupakan hasil pengumpulan data. Terdapat beberapa atribut yang dihapus seperti id user, username, coment id, profile url, avatar url, Date. Atribut yang diproses adalah kolom coment text yang berisikan komentar. Berikut merupakan hasil pengumpulan data menggunakan teknik *crawling*.

B. Pelabelan Data

Data yang diperoleh melalui teknik *crawling* dengan menghilangkan beberapa atribut yang tidak digunakan tahap selanjutnya yaitu pelabelan data. Data dibagi menjadi sentimen positif bernilai 1 dan negatif bernilai 0. Berikut data yang telah dilabeli dan divalidasi:

No	Commen Text	SENTIMEN	
0	1	Senangnyaa	1
1	2	Sehat dan bahagia selalu...Pakdhe bersama keluarga ! TUHAN TIDAK PERNAH TIDUR...untuk apa yg sel...	1
2	3	Wahhh senangnya para wartawan bisa foto sama bapak, mereka yg selalu di belakang layar tapi mere...	1
3	4	keluarga bahagia,, yg patut d contoh	1
4	5	Ini baru presiden menerima media masa untuk bersahabat dan teman udeh dech intinya pa jokowi ma...	1
...
995	996	Kok aku sedih yaðŸ˜¸..bntr lagi presidenku ga menjabat lagi, trmksh presidenku syg atas jasa BPK ...	1
996	997	Senangnya dlm hati cuan cair lagi....	1
997	998	Waktu alm ayahku sakit juga sampe kudu nangis deh, akses jalan ancur parahh bawa orang sakit kri...	0
998	999	Dulu aq sangat mengagumimu pak...Dan sangat...nSekarang kog ada rasa kecewa ya di hatiku....ðŸ˜¸	0
999	1000	Sedih kalau membayangkan Jkw mengakhiri hayatnya TIDAK DGN KHUSNUL KHATIMAH ðŸ˜¸ðŸ˜¸ðŸ˜¸	1

1000 rows x 3 columns

Gambar 5. Pelabelan Data Komentar

Pelabelan data dan validasi dilakukan oleh Bapak Agus Milu Susetyo, M.Pd., MCE selaku dosen FKIP Bahasa Indonesia Universitas Muhammadiyah Jember.

C. Text Preprocessing

Tahapan *Text Preprocessing* terbagi menjadi beberapa tahapan yaitu *cleansing*, *stopword removal*, *stemming*, *tokenizing*.

a. Cleansing

Komponen-komponen tersebut tidak memiliki pengaruh terhadap sentimen maka dibuang nantinya. Berikut ini hasil dari proses *Cleansing*:

Tabel 1. Hasil *Cleansing Text*

No	Cleaned_Text	SENTIMEN
1	Senangnyaa	1

No	Cleaned_Text	SENTIMEN
	sehat dan bahagia selalupakdhe bersama keluarga tuhan tidak pernah tiduruntuk apa yg selalu bapak perbuat bagi indonesia	1
2	perbuatan baik selalu dibalas dg kebaikantetappresiden	
	wahhh senangnya para wartawan bisa foto sama bapak mereka yg selalu di belakang layar tapi merekalah yg membawa org jd terkenal yah harus bisa bersahabat dgn mrka selama mereka memberitakan	1
3	yg tepat semangat para media	
4	keluarga bahagia yg patut d contoh	1
	ini baru presiden menerima media masa untuk bersahabat dan teman udeh dech intinya pa jokowi mantap sy sangat senang klu di	1
5	foto bareng sm semua media masa	
....
	kok aku sedih yabntr lagi presidenku ga menjabat lagitrmksh	1
996	presidenku syg atas jasa bpk slma ini u negara dan warga Indonesia	
997	senangnya dlm hati cuan cair lagi	1
	waktu alm ayahku sakit juga sampe kudu nangis deh akses jalan	0
998	ancur parahh bawa orang sakit kritis sedih banget pak	
	dulu aq sangat mengagumimu pak dan sangat	0
999	sekarang kog ada rasa kecewa ya di hatiku	
	sedih kalau membayangkan jkw mengakhiri hayatnya tidak dgn	1
1000	khusnul khatimah	

b. *Stopword Removal*

Proses ini melibatkan penghapusan kata yang tidak berpengaruh dalam menentukan suatu kategori sentimen pada komentar.

Tabel 2. Hasil *Stopword Removal*

No	<i>Stopword Removal</i>	SENTIMEN
1	Senangnyaa	1

No	<i>Stopword Removal</i>	SENTIMEN
	sehat bahagia selalupakdhe bersama keluarga tuhan pernah tiduruntuk apa yg selalu bapak perbuat indonesia perbuatan baik	
2	selalu dibalas dg kebaikantetappresiden	1
	wahhh senangnya wartawan foto sama bapak yg selalu belakang layar merekalah yg membawa org jd terkenal yah bisa bersahabat	
3	dgn mrka selama memberitakan yg tepat semangat media	1
4	keluarga bahagia yg patut d contoh	1
	baru presiden menerima media masa bersahabat teman udeh dech intinya pa jokowi mantap sy sangat senang klu foto bareng sm	
5	semua media masa	1
....
	kok aku sedih yabntr presidenku ga menjabat lagitrmksh presidenku	
996	syg atas jasa bpk slma u negara warga Indonesia	1
997	senangnya dlm hati cuan cair	1
	waktu alm ayahku sakit sampe kudu nangis deh akses jalan ancur	
998	parahh bawa orang sakit kritis sedih banget pak	0
	dulu aq sangat mengagumimu pak dan sangat	
999	sekarang kog rasa kecewa di hatiku	0
	sedih kalau membayangkan jkw mengakhiri hayatnya dgn khusnul	
1000	khatimah	1

c. Stemming

Pengubahan kata berimbuhan ke bentuk dasarnya disebut sebagai proses *stemming*. Proses ini juga merupakan salah satu langkah dalam *Information Retrieval*. Berikut ini hasil dari proses *Stemming*:

Tabel 3. Hasil *Stemming Text*

No	<i>Stemmed Text</i>	SENTIMEN
1	Senangnyaa	1

No	Stemmed_Text	SENTIMEN
	sehat bahagia selalupakdhe sama keluarga tuhan pernah tiduruntuk apa yg selalu bapak buat indonesia buat baik selalu balas dg	
2	kebaikantetappresiden	1
	wahhh senang wartawan foto sama bapak yg selalu belakang layar mereka yg bawa org jd kenal yah bisa sahabat dgn mrka lama berita	
3	yg tepat semangat media	1
4	keluarga bahagia yg patut d contoh	1
	baru presiden terima media masa sahabat teman udeh dech inti pa jokowi mantap sy sangat senang klu foto bareng sm semua media	
5	masa	1
.....
	kok aku sedih yabntr presiden ga jabat lagitrmksh presiden syg atas	
996	jasa bpk slma u negara warga Indonesia	1
997	senang dlm hati cuan cair	1
	waktu alm ayah sakit sampe kudu nang deh akses jalan ancur parahh	
998	bawa orang sakit kritis sedih banget pak	0
	dulu aq sangat kagum pak dan sangat sekarang kog rasa kecewa di	
999	hati	0
1000	sedih kalau bayang jkw akhir hayat dgn khusnul khatimah	1

d. *Tokenizing*

Pada tahap ini kata yang terdapat spasi diantara dua kata dalam satu komentar *Instagram* dipisahkan. Hasil ini dilakukan agar tahap *preprocessing* selanjutnya dapat berjalan. Berikut ini hasil proses *Tokenizing*:

Tabel 4. Hasil *Tokenizing*

No	Tokenized_Text	SENTIMEN
1	['senangnyaa']	1
	['sehat', 'bahagia', 'selalupakdhe', 'sama', 'keluarga', 'tuhan', 'pernah', 'tiduruntuk', 'apa', 'yg', 'selalu', 'bapak', 'buat', 'indonesia', 'buat',	
2	'baik', 'selalu', 'balas', 'dg', 'kebaikantetappresiden']	1

No	<i>Tokenized_Text</i>	SENTIMEN
3	['wahhh', 'senang', 'wartawan', 'foto', 'sama', 'bapak', 'yg', 'selalu', 'belakang', 'layar', 'mereka', 'yg', 'bawa', 'org', 'jd', 'kenal', 'yah', 'bisa', 'sahabat', 'dgn', 'mrka', 'lama', 'berita', 'yg', 'tepat', 'semangat', 'media']	1
4	['keluarga', 'bahagia', 'yg', 'patut', 'd', 'contoh']	1
5	['baru', 'presiden', 'terima', 'media', 'masa', 'sahabat', 'teman', 'udeh', 'dech', 'inti', 'pa', 'jokowi', 'mantap', 'sy', 'sangat', 'senang', 'klu', 'foto', 'bareng', 'sm', 'semua', 'media', 'masa']	1
....
996	['kok', 'aku', 'sedih', 'yabntr', 'presiden', 'ga', 'jabat', 'lagitrmksh', 'presiden', 'syg', 'atas', 'jasa', 'bpk', 'slma', 'u', 'negara', 'warga', 'indonesia']	1
997	['senang', 'dlm', 'hati', 'cuan', 'cair']	1
998	['waktu', 'alm', 'ayah', 'sakit', 'sampe', 'kudu', 'nang', 'deh', 'akses', 'jalan', 'ancur', 'parahh', 'bawa', 'orang', 'sakit', 'kritis', 'sedih', 'banget', 'pak']	0
999	['dulu', 'aq', 'sangat', 'kagum', 'pak', 'dan', 'sangat', 'sekarang', 'kog', 'rasa', 'kecewa', 'di', 'hati']	0
1000	['sedih', 'kalau', 'bayang', 'jkw', 'akhir', 'hayat', 'dgn', 'khusnul', 'khatimah']	1

D. Implementasi Metode

Setelah melewati tahap *Text Preprocessing*, tahap selanjutnya yaitu melakukan pembobotan kata pada setiap term yang terdapat pada komentar. Proses pembobotan kata pada penelitian ini menggunakan pembobotan kata *TF-IDF*. Berikut ini merupakan contoh pembobotan kata berdasarkan data positif dan negatif:

Tabel 5. Hasil Pembobotan Kata menggunakan TF-IDF

No	<i>Tokenized_Text</i>	<i>TF-IDF</i>
1	['senangnyaa']	['1.6094']
2	['buat', 'selalu', 'sehat', 'dg', 'tiduruntuk', 'pernah', 'kebaikantetappresiden', 'baik', 'tuhan', 'indonesia']	['0.0804', '0.458', '0.0916', '0.0458', '0.0458', '0.0804', '0.0804', '0.0804', '0.0804']

No	Tokenized_Text	TF-IDF
3	['dgn', 'org', 'layar', 'mereka', 'mrka', 'jd', 'bisa', 'semangat', 'berita', 'lama']	['0.0596', '0.0596', '0.0596', '0.0596', '0.0596', '0.0596', '0.0596', '0.0596', '0.0596', '0.0596',]
4	['patut', 'contoh', 'bahagia', 'keluarga', 'yg', 'masa', 'bisa', 'sy', 'kebaikantetappresiden']	['0.2682', 0.2682, '0.1527', '0.1527', '0.0851']
5	['masa', 'media', 'pa', 'inti', 'dech', 'terima', 'bareng', 'sy', 'jokowi', 'semua']	['0.1399', '0.0796', '0.0699', '0.0699', '0.0699', '0.0699', '0.0699', '0.0699', '0.0699', '0.0699']
....
996	['kok', 'aku', 'sedih', 'yabntr', 'presiden', 'ga', 'jabat', 'lagitrmksh', 'presiden', 'syg', 'atas', 'jasa', 'bpk', 'slma', 'u', 'negara', 'warga', 'indonesia']	['0.1788', '0.0894', 0.0894', 0.0894', 0.0894', 0.0894', 0.0894', 0.0894', 0.0894', 0.0894',]
997	['senang', 'dlm', 'hati', 'cuan', 'cair']	['0.3218', '0.3218', '0.3218', '0.3218', '0.1832']
998	['waktu', 'alm', 'ayah', 'sakit', 'sampe', 'kudu', 'nang', 'deh', 'akses', 'jalan', 'ancur', 'parahh', 'bawa', 'orang', 'sakit', 'kritis', 'sedih', 'banget', 'pak']	['0.1694', '0.0847', '0.0847', '0.0847', '0.0847', '0.0847', '0.0847', '0.0847', '0.0847', '0.0847', '0.0847']
999	['dulu', 'aq', 'sangat', 'kagum', 'pak', 'dan', 'sangat', 'sekarang', 'kog', 'rasa', 'kecewa', 'di', 'hati']	['0.2476', '0.1238', '0.1238', '0.1238', '0.1238', '0.1238', '0.1238', '0.1238', '0.1238', '0.1238']
1000	['sedih', 'kalau', 'bayang', 'jkw', 'akhir', 'hayat', 'dgn', 'khusnul', 'khatimah']	['0.1788', '0.1788', '0.1788', '0.1788', '0.1788', '0.1788', '0.1788', '0.1788', '0.1788', '0.0567']

E. Pembagian dan pengujian Datase

Proses inti yaitu klasifikasi menggunakan metode *Naïve Bayes Classifier*. Berikut adalah hasil pengujian klasifikasi data dengan menggunakan *K-Fold Cross Validation*:

Tabel 6. Hasil Klasifikasi *Naive Bayes Classifier*

Langkah Uji	<i>Naïve Bayes Classifier</i>
	Akurasi
Uji Coba 1	81,5%
Uji Coba 2	79%
Uji Coba 3	78,5%
Uji Coba 4	80%
Uji Coba 5	82,5%
Akurasi Rata-Rata	80,3%

Berdasarkan tabel diatas diperoleh hasil penggunaan metode *Naïve Bayes Classifier* dengan rata-rata akurasi sebesar 80.3%, dapat disimpulkan bahwa model memiliki performa yang cukup baik dalam memprediksi label kelas pada data yang tidak terlihat sebelumnya.

F. Evaluasi Hasil *Naïve Bayes classifier*

Tahap terakhir dalam penelitian ini adalah melakukan evaluasi hasil dari *naïve bayes classifier* dengan menggunakan pengukuran. Pengukuran merupakan penentuan kriteria menggunakan *Confusion Matrix* dan pengukuran tingkat akurasi, presisi serta *recall*.

Tabel 7. Hasil Sentimen data Actual dan Prediksi

Dokumen	Nilai Aktual	Nilai Prediksi	Hasil
Dokumen 1	1	1	TP
Dokumen 2	1	1	TP
Dokumen 3	1	1	TP
Dokumen 4	1	1	TP
Dokumen 5	1	1	TP
Dokumen 6	1	0	FN
Dokumen 7	1	1	TP
Dokumen 8	1	1	TP
Dokumen 9	1	1	TP

Dokumen	Nilai Aktual	Nilai Prediksi	Hasil
Dokumen 10	1	0	FN
Dokumen 11	1	1	TP
Dokumen 12	1	0	FN
Dokumen 13	0	0	TN
.....
Dokumen 988	0	1	FP
Dokumen 989	1	0	FN
Dokumen 990	1	1	TP
Dokumen 991	1	1	TP
Dokumen 992	1	1	TP
Dokumen 993	1	1	TP
Dokumen 994	1	1	TP
Dokumen 995	1	1	TP
Dokumen 996	1	1	TP
Dokumen 997	1	1	TP
Dokumen 998	0	0	TN
Dokumen 999	0	0	TN
Dokumen 1000	1	0	FN

Tabel 1 Hasil Confusion Matrix

	<i>Predicted Class</i>	
	Positif (1)	Negatif (0)
<i>Actual Class</i>	Positif (1)	TP=635 FP=107
	Negatif (0)	FN=52 TN=206

Dari pengukuran diatas dapat disimpulkan bahwasannya didapatkan akurasi sebesar 84,1%. Akurasi mengukur sejauh mana yang dapat memprediksi dengan baik kelas positif maupun negatif. Presisi memiliki presentase sebesar 85,5% mengukur sejauh mana prediksi positif yang dilakukan oleh kinerja yang benar-benar relevan. Dan recall memiliki presentase sebesar 92,4%, yang mengukur sejauh mana dapat mendeteksi atau mengingat hasil positif yang sebenarnya.

KESIMPULAN DAN SARAN

Hasil yang didapat dari dataset sejumlah 1000 yang dibagi menjadi 5 bagian yang seimbang, dengan 800 data latih (training) dan 200 data uji (testing) sehingga menghasilkan akurasi rata-rata 80,3%. Dengan Komentar positif sejumlah 687 komentar, dan komentar negatif sejumlah 313 komentar. Akurasi yang tinggi, menyatakan bahwa metode Naïve Bayes Classifier dapat membantu dalam mengidentifikasi sentimen yang terkandung dalam teks, baik itu sentimen negatif maupun positif. Dan dilakukan pengukuran menggunakan Confusion Matrix didapatkan akurasi sebesar 84,1%. Akurasi mengukur sejauh mana yang dapat memprediksi dengan baik kelas positif maupun negatif. Presisi memiliki presentase sebesar 85,5% mengukur sejauh mana prediksi positif yang dilakukan oleh kinerja yang benar-benar relevan. Dan recall memiliki presentase sebesar 92,4%, yang mengukur sejauh mana dapat mendeteksi atau mengingat hasil positif yang sebenarnya.

DAFTAR REFERENSI

- Azhari, M., Situmorang, Z., & Rosnelly, R. (2021). Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes. *Jurnal Media Informatika Budidarma*, 5(2), 640. <https://doi.org/10.30865/mib.v5i2.2937>
- Firdaus, A., & Firdaus, W. I. (2021). *Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan)*. 13(1), 66–78.
- Isnain, A. R., Sakti, A. I., Alita, D., & Marga, N. S. (2021). *Jakarta Menggunakan Algoritma SVM Media sosial menjadikan masyarakat mengalami pergeseran perilaku baik budaya , etika dan norma yang ada , sehingga mereka dapat mengeluarkan opini - opini yang mereka miliki . Opini merupakan suatu pendapat dari pemikiran*. 2(1), 31–37.
- Maulidina, M. K. (2020). *Analisis Sentimen Komentar Warganet Terhadap Postingan Instagram Menggunakan Metode*.
- Nisa, A. L. F. (2023). *Optimizing Arabic Language Proficiency Through Instagram Social*. 01(03), 115–124. <https://doi.org/10.18860/kitaba.v1i3.23415>
- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 5(2), 697–711.
- Rusdaman, D., & Rosiyadi, D. (2019). *Metode Naïve Bayes Classifier Dan Support Vector Machine*. 4(2), 230–235.
- Sari, H., Ginting, G. L., Zebua, T., & Mesran. (2021). Penerapan Algoritma Text Mining dan

TF-IDF untuk Pengelompokan Topik Skripsi pada Aplikasi Repository STMIK Budi Darma. *TIN: Terapan Informatika Nusantara*, 2(7), 414–432.

Tempola, F., Muhammad, M., & Khairan, A. (2018). Perbandingan Klasifikasi Antara Knn Dan Naive Bayes Pada Penentuan Status Gunung Berapi Dengan K-Fold Cross Validation Comparison of Classification Between Knn and Naive Bayes At the Determination of the Volcanic Status With K-Fold Cross Validation. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, 5(5), 577–584. <https://doi.org/10.25126/jtiik20185983>

Thoyyibah, Kurniawan, F., & Taryo, T. (2024). *Dasar - dasar Machine Learning Pada Google Colabs*. Eureka Media Aksara.

Watratan, A. F., Puspita, A., & Moeis, D. (2020). *Journal Of Applied Computer Science And Technology (JACOST) Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia*. 1(1), 7–14.

Widayat, W. (2021). *Analisis Sentimen Movie Review menggunakan Word2Vec dan metode LSTM Deep Learning*. 5(2014), 1018–1026. <https://doi.org/10.30865/mib.v5i3.3111>

Widowati, T. T., & Sadikin, M. (2021). Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 11(2), 626–636. <https://doi.org/10.24176/simet.v11i2.4568>

Zhafira, D. F., Rahayudi, B., & Indriati. (2021). Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naive Bayes dan Pembobotan TF-IDF Berdasarkan Komentar pada Youtube. *Jurnal Sistem Informasi, Teknologi Informasi, Dan Edukasi Sistem Informasi*, 2(1), 55–63. <https://doi.org/10.25126/justsi.v2i1.24>