

ANALISIS SENTIMEN MASYARAKAT TERHADAP ISU IJAZAH PALSU JOKOWI MENGGUNAKAN ALORITMA NAÏVE BAYES PADA PLATFROM X (TWITTER)

Sandy Hafizul Pratama¹, Marsella Rukman², Putri S.A Tallo³, Imelda⁴
sandyexpressiastudio@gmail.com¹, marsellarukman360@gmail.com², ptrytallo@gmail.com³,
imeli7057@gmail.com⁴
Universitas Papua

ABSTRAK

Perkembangan media sosial, khususnya platform X (Twitter), telah menjadi sarana utama bagi masyarakat dalam menyampaikan opini terhadap berbagai isu publik, termasuk isu ijazah palsu yang dikaitkan dengan mantan Presiden ke 7 Joko Widodo. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap isu tersebut menggunakan algoritma Naïve Bayes. Data yang digunakan merupakan data sekunder yang diperoleh melalui proses crawling dari platform X (Twitter) dengan kata kunci yang relevan. Tahap pengolahan data meliputi preprocessing yang terdiri dari cleaning, case folding, tokenizing, stopword removal, dan stemming. Selanjutnya, data ditransformasikan menggunakan metode TF-IDF sebelum dilakukan proses klasifikasi ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral menggunakan algoritma Naïve Bayes. Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes mampu mengklasifikasikan sentimen dengan tingkat akurasi yang cukup baik. Selain itu, hasil analisis memberikan gambaran mengenai kecenderungan opini publik terhadap isu ijazah yang berkembang di media sosial. Dengan demikian, penelitian ini dapat menjadi referensi dalam pengembangan analisis sentimen serta membantu memahami persepsi masyarakat terhadap suatu isu secara lebih objektif.

Kata Kunci: Analisis Sentimen; Naïve Bayes; Twitter; Ijazah Jokowi; Text Mining; Media Sosial

ABSTRACT

The development of social media, especially platform X (Twitter), has become the primary means for people to express their opinions on various public issues, including the issue of fake diplomas associated with the former 7th President Joko Widodo. This study aims to analyze public sentiment towards this issue using the Naïve Bayes algorithm. The data used is secondary data obtained through a crawling process from platform X (Twitter) with relevant keywords. The data processing stage includes preprocessing consisting of cleaning, case folding, tokenizing, stopword removal, and stemming. Next, the data is transformed using the TF-IDF method before being classified into three sentiment categories, namely positive, negative, and neutral using the Naïve Bayes algorithm. The results of the study show that the Naïve Bayes algorithm is able to classify sentiment with a fairly good level of accuracy. In addition, the results of the analysis provide an overview of the tendency of public opinion towards the issue of diplomas that develops on social media. Thus, this study can be a reference in the development of sentiment analysis and help understand public perception of an issue more objectively.

Keywords: Sentiment Analysis; Naïve Bayes; Twitter; Jokowi Diploma Issue; Text Mining; Social Media

PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi yang pesat telah mendorong peningkatan penggunaan media sosial sebagai sarana komunikasi dan penyebaran informasi. Salah satu platform yang banyak digunakan oleh masyarakat adalah X (Twitter), yang memungkinkan pengguna untuk menyampaikan opini secara cepat dan luas terhadap berbagai isu yang berkembang. Media sosial tidak hanya berfungsi sebagai media berbagi

informasi, tetapi juga menjadi ruang publik digital yang mencerminkan persepsi dan sentimen masyarakat terhadap suatu peristiwa atau isu tertentu (Hidayat et al., 2024).

Salah satu isu yang sempat menjadi perhatian publik adalah dugaan ijazah palsu yang dikaitkan dengan mantan Presiden Joko Widodo. Isu ini menimbulkan beragam respons dari masyarakat, mulai dari opini yang bersifat positif, negatif, hingga netral. Banyaknya opini yang tersebar di media sosial menyebabkan kesulitan dalam memahami kecenderungan persepsi publik secara manual. Oleh karena itu, diperlukan suatu metode yang mampu mengolah data teks dalam jumlah besar secara otomatis untuk mengidentifikasi sentimen masyarakat terhadap isu tersebut.

Analisis sentimen merupakan salah satu cabang dari data mining dan Natural Language Processing (NLP) yang digunakan untuk mengidentifikasi serta mengklasifikasikan opini dalam bentuk teks ke dalam kategori tertentu, seperti positif, negatif, dan netral (Nugroho et al., 2023). Berbagai penelitian sebelumnya telah menerapkan analisis sentimen pada data media sosial, khususnya Twitter, untuk mengkaji opini publik terhadap berbagai topik seperti kebijakan pemerintah, layanan digital, maupun isu sosial (Fitrianti & Yudhistira, 2023). Salah satu metode yang sering digunakan dalam klasifikasi teks adalah algoritma Naïve Bayes, yang dikenal karena kesederhanaannya, efisiensi komputasi, serta kemampuannya dalam menangani data teks dengan cukup baik (Yudistira & Isnain, 2024).

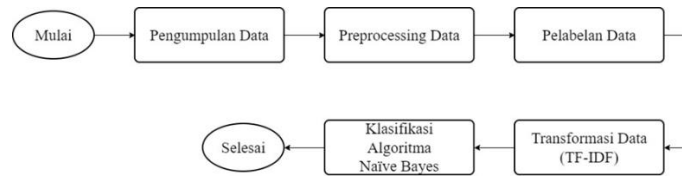
Sejumlah penelitian terdahulu menunjukkan bahwa algoritma Naïve Bayes mampu menghasilkan tingkat akurasi yang baik dalam analisis sentimen, khususnya pada data berbasis teks pendek seperti tweet. Selain itu, penerapan teknik preprocessing seperti cleaning, tokenizing, stopword removal, dan stemming terbukti dapat meningkatkan performa klasifikasi (Hidayat et al., 2024). Namun demikian, sebagian besar penelitian masih berfokus pada topik umum, seperti layanan aplikasi, produk digital, atau kebijakan tertentu, dan belum banyak yang secara khusus mengkaji isu sensitif yang berkaitan dengan figur publik serta persepsi masyarakat secara real-time di media sosial.

Berdasarkan kondisi tersebut, terdapat kesenjangan penelitian (gap analysis), yaitu masih terbatasnya studi yang secara spesifik menganalisis sentimen masyarakat terhadap isu ijazah palsu yang dikaitkan dengan mantan Presiden Joko Widodo menggunakan data dari platform X (Twitter). Selain itu, diperlukan penelitian yang mampu memberikan gambaran yang lebih aktual mengenai persepsi publik terhadap isu yang berkembang secara cepat di media sosial. Kebaruan (novelty) dalam penelitian ini terletak pada fokus kajian terhadap isu yang sedang berkembang serta pemanfaatan algoritma Naïve Bayes untuk mengklasifikasikan sentimen masyarakat berdasarkan data terbaru dari media sosial.

Berdasarkan latar belakang dan kesenjangan penelitian tersebut, tujuan penelitian ini adalah untuk menganalisis sentimen masyarakat terhadap isu ijazah palsu yang dikaitkan dengan mantan Presiden Joko Widodo di platform X (Twitter) menggunakan algoritma Naïve Bayes. Selain itu, penelitian ini juga bertujuan untuk mengukur kinerja algoritma dalam mengklasifikasikan sentimen serta memberikan gambaran mengenai kecenderungan opini publik terhadap isu tersebut.

METODOLOGI PENELITIAN

Untuk memberikan gambaran alur penelitian secara sistematis, disajikan diagram metode penelitian yang menunjukkan tahapan utama mulai dari proses pengumpulan data hingga klasifikasi menggunakan algoritma Naïve Bayes. Diagram ini digunakan untuk memperjelas langkah-langkah dalam pengolahan dan analisis data sentimen masyarakat terhadap isu ijazah mantan Presiden Joko Widodo pada platform X (Twitter).



Gambar 1. Diagram Sistematis Pengolahan Data

Sumber : Peneliti

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap isu ijazah palsu yang dikaitkan dengan mantan Presiden Joko Widodo berdasarkan opini yang disampaikan pada platform X (Twitter). Untuk mencapai tujuan tersebut, digunakan pendekatan machine learning dengan algoritma Naïve Bayes sebagai metode klasifikasi sentimen. Algoritma ini dipilih karena memiliki efisiensi yang tinggi dalam pengolahan data teks serta mampu memberikan performa yang baik pada data berukuran besar dan tidak terstruktur (Yudistira & Isnain, 2024).

Tahap pertama dalam penelitian ini adalah pengumpulan data (crawling data), yang dilakukan menggunakan tools tweet-harvest dengan memanfaatkan kata kunci yang relevan dengan isu ijazah mantan Presiden Joko Widodo. Parameter yang digunakan meliputi kata kunci pencarian, rentang waktu pengambilan data, serta jumlah data yang akan dikumpulkan. Data yang diperoleh berupa tweet dalam bahasa Indonesia yang mengandung opini masyarakat terkait isu yang diteliti. Teknik crawling banyak digunakan dalam penelitian analisis sentimen karena mampu mengumpulkan data secara otomatis dalam jumlah besar (Hidayat et al., 2024).

Data yang telah diperoleh kemudian diproses melalui tahapan preprocessing untuk meningkatkan kualitas data sebelum dianalisis. Tahapan ini meliputi cleaning, case folding, tokenizing, stopword removal, dan stemming. Proses ini bertujuan untuk menghilangkan noise serta menyeragamkan bentuk teks sehingga lebih mudah dianalisis. Tahapan preprocessing terbukti berperan penting dalam meningkatkan akurasi model klasifikasi teks (Fitrianti & Yudhistira, 2023).

Tahap berikutnya adalah pelabelan data (labelling), yaitu proses pemberian label sentimen pada setiap tweet. Pelabelan dilakukan secara otomatis menggunakan metode VADER (Valence Aware Dictionary and Sentiment Reasoner), yang mampu mengidentifikasi polaritas sentimen berdasarkan skor yang dihasilkan. Data kemudian dikategorikan ke dalam tiga kelas sentimen, yaitu positif, negatif, dan netral. Pendekatan ini umum digunakan dalam analisis sentimen untuk mempercepat proses anotasi data (Nugroho et al., 2023).

Setelah proses pelabelan, dilakukan tahap transformasi data menggunakan metode TF-IDF (Term Frequency–Inverse Document Frequency). Metode ini digunakan untuk mengubah data teks menjadi representasi numerik berdasarkan bobot kemunculan kata dalam dokumen, sehingga dapat digunakan sebagai input dalam proses klasifikasi. Penggunaan TF-IDF telah terbukti efektif dalam merepresentasikan fitur teks pada berbagai penelitian analisis sentimen (Hidayat et al., 2024).

Selanjutnya, dilakukan proses klasifikasi menggunakan algoritma Naïve Bayes. Algoritma ini bekerja dengan menghitung probabilitas kemunculan kata dalam setiap kelas sentimen, kemudian menentukan kelas dengan probabilitas tertinggi sebagai hasil klasifikasi. Dataset dibagi menjadi data pelatihan sebesar 80% dan data pengujian sebesar 20% untuk membangun dan menguji model. Metode Naïve Bayes banyak digunakan karena kesederhanaan dan efisiensinya dalam klasifikasi teks (Yudistira & Isnain, 2024).

HASIL PEMBAHASAN

Pengumpulan Data

Pada penelitian ini, tahap pertama dimulai dengan pengumpulan data yang dilakukan menggunakan teknik web crawling untuk mengambil data dari platform X (Twitter). Proses pengambilan data dilakukan dengan memanfaatkan tools tweet-harvest serta menggunakan kata kunci yang relevan dengan isu ijazah palsu yang dikaitkan dengan mantan Presiden Joko Widodo.

Data yang dikumpulkan berupa tweet dalam bahasa Indonesia yang mengandung opini masyarakat terkait isu tersebut. Proses crawling dilakukan dengan menentukan parameter seperti kata kunci pencarian, rentang waktu pengambilan data, serta jumlah data yang diambil. Hasil dari proses crawling kemudian disimpan dalam format CSV untuk memudahkan proses pengolahan data selanjutnya. Jumlah data yang berhasil dikumpulkan sebanyak ±150 tweet yang relevan dengan topik penelitian. Data tersebut selanjutnya digunakan sebagai bahan utama dalam proses analisis sentimen. Contoh data yang dihasilkan dari proses crawling dapat dilihat pada tabel berikut.

Gambar 2. Hasil Crawling Data

conversation_id	created_at	source	full_text	lang	media_url	reply_to_screen_name	in_reply_to_screen_name
2.04544E+18	18 Sat Apr 18 12:16:02 +0000 2020	4	@LambeSahajia Masih aja ributin ijazah palsu kenapa gak dari dulu elabih... udah jadi pria si	2.0455E+18			LambeSahajia
2.04538E+18	18 Sat Apr 18 12:15:53 +0000 2020	5	@DokterTifa @MuhammadAnifP09 @RKMTRoy2020 yg awalnya gak yakin Pak Iku ijazah pal	2.0455E+18			DokterTifa
2.04546E+18	18 Sat Apr 18 12:09:47 +0000 2020	0	@NaniMolica Mungkin Allah sudah bosan dengan perkara Ijazah Palsu ini terlalu besar mudh	2.0455E+18	https://pbs.twimg.com/media/HGLRjUcAEPOz.jpg		NaniMolica
2.04544E+18	18 Sat Apr 18 12:09:39 +0000 2020	7	@LambeSahajia Ika Anda pada Posisi Pak Jokowi apa Yg Anda Lakukan..?? Fucking Bull SHIT	2.0455E+18			LambeSahajia
2.04496E+18	18 Sat Apr 18 12:08:57 +0000 2020	0	@Lukesulio @DokterTifa @Rizantaza Bahkan saki ahi sekarang lagi main ijazah palsu m	2.0455E+18			Lukesulio
2.04532E+18	18 Sat Apr 18 11:57:15 +0000 2020	0	@BambangMulyono SESAMA IJAZAH PALSU HARUS SALING MENDUKUNG...	2.0455E+18			BambangMulyono
2.04547E+18	18 Sat Apr 18 11:52:47 +0000 2020	0	terima kasih banget Ijazah simon palsu km warnanya putih padahal harusnya kuning wh si pr	2.0455E+18			Lukesulio
2.04547E+18	18 Sat Apr 18 11:52:20 +0000 2020	0	kegiatan teroris yang sng yg nganggap ky teroris km ky tanggabin sampe jadi ky akhm	2.0455E+18			sekolah_amb
2.04538E+18	18 Sat Apr 18 11:51:59 +0000 2020	0	@DokterTifa @MuhammadAnifP09 @RKMTRoy2020 itu dokter apa dan keluarga dari Madan?	2.0455E+18			DokterTifa
2.04475E+18	18 Sat Apr 18 11:31:46 +0000 2020	0	@Rizantaza Apa sih yg elu tahu dan paham selain Menjilat Masih percaya pendukung Ijazah P	2.0455E+18			equitavee
2.04546E+18	18 Sat Apr 18 11:28:28 +0000 2020	0	AllahuMahaMaju terus Perjuangn bngtur Ijazah Palsu kang Pembolong	2.0455E+18			LambeSahajia
2.04546E+18	18 Sat Apr 18 11:27:46 +0000 2020	21	@LambeSahajia kakak @Pak_Iku ijahe tau selo jd pangeran opesy dngan ke KPU baw	2.0455E+18			LambeSahajia
2.04538E+18	18 Sat Apr 18 11:21:32 +0000 2020	0	@DokterTifa @RKMTRoy2020 saya percaya ijazah si jak palsu 100% jangan takut rakyat baru	2.0455E+18			DokterTifa
2.04546E+18	18 Sat Apr 18 11:20:28 +0000 2020	0	Padahal orang dibelakang layar dalam hal ini... Lu sama majikan gue samal Ijazah Palsu	2.0455E+18			DokterTifa
2.04532E+18	18 Sat Apr 18 11:16:47 +0000 2020	0	@Kerjastasi_id Gara2 Ijazah Palsu @ Jokowi Raja Ngbul https://t.co/gTWwQdP	2.0455E+18	https://pbs.twimg.com/media/HGLub21asAEQMBB.jpg?kerjastasi_id		kerjastasi_id
2.04546E+18	18 Sat Apr 18 11:15:55 +0000 2020	0	waga negara kalo ngelamar kerja hrs ada SKCK .. lusan apa... kerja dibatasin umur... nah bar	2.0455E+18			LambeSahajia
2.04544E+18	18 Sat Apr 18 12:16:02 +0000 2020	4	@LambeSahajia Masih aja ributin ijazah palsu kenapa gak dari dulu elabih... udah jadi pria si	2.0455E+18			LambeSahajia
2.04538E+18	18 Sat Apr 18 12:15:53 +0000 2020	5	@DokterTifa @MuhammadAnifP09 @RKMTRoy2020 yg awalnya gak yakin Pak Iku ijazah pal	2.0455E+18			DokterTifa
2.04546E+18	18 Sat Apr 18 12:09:47 +0000 2020	0	@NaniMolica Mungkin Allah sudah bosan dengan perkara Ijazah Palsu ini terlalu besar mudh	2.0455E+18	https://pbs.twimg.com/media/HGLRjUcAEPOz.jpg		NaniMolica
2.04544E+18	18 Sat Apr 18 12:09:39 +0000 2020	7	@LambeSahajia Ika Anda pada Posisi Pak Jokowi apa Yg Anda Lakukan..?? Fucking Bull SHIT	2.0455E+18			LambeSahajia
2.04496E+18	18 Sat Apr 18 12:08:57 +0000 2020	0	@Lukesulio @DokterTifa @Rizantaza Bahkan saki ahi sekarang lagi main ijazah palsu m	2.0455E+18			Lukesulio
2.04532E+18	18 Sat Apr 18 11:57:15 +0000 2020	0	@BambangMulyono SESAMA IJAZAH PALSU HARUS SALING MENDUKUNG...	2.0455E+18			BambangMulyono

Sumber : Data Diolah Peneliti

Preprocessing Data

Setelah data berhasil dikumpulkan, langkah selanjutnya adalah melakukan preprocessing data. Proses ini sangat penting untuk menyiapkan data mentah agar dapat digunakan secara efektif dalam analisis maupun pembuatan model machine learning. Data yang berasal dari platform X (Twitter) umumnya masih mengandung banyak noise, seperti penggunaan bahasa tidak baku, simbol, emoji, serta elemen lain yang tidak relevan, sehingga perlu dilakukan pembersihan terlebih dahulu.

Tabel 1. Case Folding

Sebelum	Sesudah
SESAMA IJAZAH PALSU HARUS SALING MENDUKUNG...	sesama ijazah palsu harus saling mendukung...

Sumber : Data Diolah Peneliti

Tahap pertama yang dilakukan adalah case folding, yaitu proses mengubah seluruh teks menjadi huruf kecil. Tujuan dari proses ini adalah untuk menjaga konsistensi data, karena perbedaan huruf kapital dan huruf kecil tidak memiliki makna yang berbeda dalam analisis sentimen. Sebagai contoh, kata "IJAZAH", "Ijazah" akan dianggap sebagai kata yang sama setelah melalui proses ini.

Tabel 2. Cleaning Data

Sebelum	Sesudah
apa sih yg elu tahu dan paham selain menjilat masih percaya pendukung ijazah palsu ?	apa sih yg elu tahu dan paham selain menjilat masih percaya pendukung ijazah palsu

Sumber : Data Diolah Peneliti

Selanjutnya dilakukan proses cleaning, yaitu penghapusan elemen-elemen yang tidak relevan seperti URL, mention (@username), hashtag (#topik), tanda baca, angka, serta karakter khusus lainnya. Proses ini bertujuan untuk mengurangi noise dalam data sehingga hanya menyisakan teks yang benar-benar memiliki makna.

Tabel 3. Tokenisasi

Sebelum	Sesudah
alhamdulillah maju terus perjuangan bongkar ijazah palsu sang pembohong	alhamdulillah, maju, terus, perjuangan, bongkar, ijazah, palsu, sang, pembohong

Sumber : Data Diolah Peneliti

Tahap berikutnya adalah tokenizing, yaitu proses memecah teks menjadi unit-unit kata atau token. Proses ini diperlukan agar setiap kata dapat dianalisis secara terpisah dalam tahap selanjutnya. Setelah itu, dilakukan stopword removal, yaitu penghapusan kata-kata umum yang sering muncul tetapi tidak memiliki makna penting dalam analisis, seperti “dan”, “yang”, “di”, dan “ke”.

Tabel 4. Stopword Removal

Sebelum	Sesudah
ibu dokter saya dan keluarga dari medan sumatera utara ikut mendukung ibu dokter dan mas roy suryo kita pun haqqul yakin ijazah jokowi adalah palsu	ibu, dokter, keluarga, medan, sumatera, utara, mendukung, ibu, dokter, mas, roy, suryo, haqqul, yakin, ijazah, jokowi, palsu

Sumber : Data Diolah Peneliti

Tahapan terakhir adalah stemming, yaitu proses mengubah kata ke bentuk dasarnya. Misalnya, kata “mendukung”, “didukung”, dan “dukungan” akan diubah menjadi “dukung”. Proses ini bertujuan untuk menyederhanakan variasi kata sehingga dapat meningkatkan akurasi model dalam mengenali pola sentimen.

Dengan melalui tahapan preprocessing tersebut, data teks yang semula tidak terstruktur menjadi lebih bersih, terstruktur, dan siap digunakan dalam proses klasifikasi sentimen menggunakan algoritma Naïve Bayes.

Tabel 5. Stemming

Sebelum	Sesudah
apa sih yang kamu tahu dan paham selain menjilat masih percaya pendukung ijazah palsu	tahu, paham, jilat, percaya, dukung, ijazah, palsu

Sumber : Data Diolah Peneliti

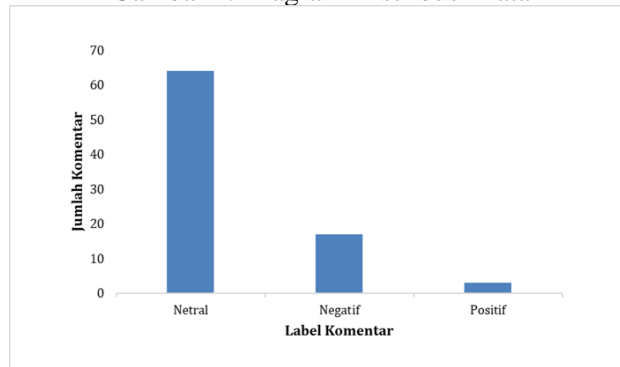
Pelabelan Data

Setelah tahap preprocessing selesai, data yang telah diproses kemudian masuk ke tahap pelabelan (labelling). Pada tahap ini, setiap tweet diberikan label sentimen untuk mengklasifikasikan opini masyarakat terhadap isu ijazah palsu yang dikaitkan dengan Presiden Joko Widodo.

Data dikategorikan ke dalam tiga label sentimen, yaitu positif, negatif, dan netral. Label positif diberikan pada tweet yang menunjukkan dukungan atau opini yang bersifat membela, label negatif diberikan pada tweet yang mengandung kritik, tuduhan, atau sentimen tidak mendukung, sedangkan label netral diberikan pada tweet yang bersifat informatif atau tidak menunjukkan kecenderungan opini tertentu.

Hasil pelabelan data kemudian divisualisasikan dalam bentuk diagram batang untuk menunjukkan distribusi jumlah tweet pada masing-masing kategori sentimen. Diagram tersebut menggambarkan perbandingan jumlah data pada setiap kelas, sehingga dapat diketahui kecenderungan opini publik terhadap isu yang sedang diteliti.

Gambar 2. Diagram Distribusi Data



Sumber : Data Diolah Peneliti

Berdasarkan hasil visualisasi, terlihat bahwa kategori sentimen netral memiliki jumlah paling dominan dibandingkan kategori lainnya, yaitu sekitar 60-an komentar. Selanjutnya diikuti oleh kategori negatif dengan jumlah sekitar belasan komentar. Sementara itu, kategori positif memiliki jumlah paling sedikit, hanya beberapa komentar saja. Distribusi ini menunjukkan bahwa mayoritas opini masyarakat cenderung bersifat netral terhadap isu yang berkembang di media sosial, dengan sebagian kecil menunjukkan sentimen negatif dan sangat sedikit yang memberikan respons positif.

Transformasi TF-IDF

Setelah tahap pelabelan data selesai, data yang telah memiliki label sentimen kemudian masuk ke tahap transformasi data menggunakan metode TF-IDF (Term Frequency–Inverse Document Frequency). Pada tahap ini, setiap tweet yang telah melalui proses preprocessing diubah ke dalam bentuk representasi numerik agar dapat diproses oleh algoritma machine learning.

Melalui proses ini, data teks yang semula tidak terstruktur diubah menjadi matriks fitur numerik yang dapat digunakan sebagai input dalam proses klasifikasi menggunakan algoritma Naïve Bayes. Dengan demikian, transformasi TF-IDF berperan penting dalam meningkatkan kualitas representasi data sehingga dapat membantu model dalam mengenali pola sentimen secara lebih akurat.

Gambar 3. Hasil Transformasi TF-IDF

```

***
      gak      ijazah      jokowi      malah      mereka      palsu      rismon      terus \
0 0.615992 0.101156 0.220981 0.270094 0.0 0.214578 0.0 0.270094
1 0.888539 0.145913 0.318755 0.000000 0.0 0.154759 0.0 0.000000
2 0.000000 0.686007 0.000000 0.000000 0.0 0.727595 0.0 0.000000
3 0.000000 0.230290 0.503080 0.000000 0.0 0.244251 0.0 0.000000
4 0.000000 0.686007 0.000000 0.000000 0.0 0.727595 0.0 0.000000

      udah      yg
0 0.582195 0.174907
1 0.000000 0.252295
2 0.000000 0.000000
3 0.000000 0.796378
4 0.000000 0.000000
    
```

Sumber : Data Diolah Peneliti

Berdasarkan hasil yang diperoleh, terlihat bahwa kata-kata seperti “ijazah”, “jokowi”, dan “palsu” memiliki bobot yang cukup tinggi pada beberapa dokumen, yang menunjukkan bahwa kata-kata tersebut menjadi fitur utama dalam analisis sentimen terhadap isu yang diteliti.

Klasifikasi Algoritma Naive Bayes

Setelah tahap transformasi data menggunakan TF-IDF selesai, data yang telah berbentuk representasi numerik kemudian digunakan dalam proses klasifikasi dengan algoritma Naïve Bayes. Pada tahap ini, model akan mempelajari pola distribusi kata dalam setiap kelas sentimen dan mengklasifikasikan setiap tweet ke dalam kategori positif, negatif, atau netral berdasarkan probabilitas tertinggi.

Gambar 4. Akurasi model

Akurasi Model: 82.35%				
Laporan Klasifikasi:				
	precision	recall	f1-score	support
negatif	1.00	0.33	0.50	3
netral	0.81	1.00	0.90	13
positif	0.00	0.00	0.00	1
accuracy			0.82	17
macro avg	0.60	0.44	0.47	17
weighted avg	0.80	0.82	0.77	17

Sumber : Data Diolah Peneliti

Berdasarkan hasil evaluasi menggunakan confusion matrix dan classification report, diperoleh nilai akurasi model sebesar 82.35%. Hal ini menunjukkan bahwa model Naïve Bayes mampu mengklasifikasikan sebagian besar data uji dengan cukup baik.

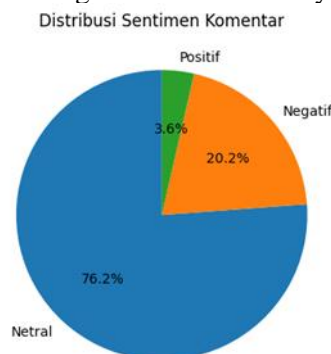
Jika dilihat berdasarkan masing-masing kelas, model menunjukkan performa yang berbeda. Pada kelas netral, model memiliki nilai precision sebesar 0.81, recall sebesar 1.00, dan f1-score sebesar 0.90, yang menunjukkan bahwa model sangat baik dalam mengenali data dengan sentimen netral.

Pada kelas negatif, model memiliki precision sebesar 1.00 namun recall hanya sebesar 0.33. Hal ini menunjukkan bahwa meskipun prediksi negatif yang dihasilkan cenderung tepat, model masih belum mampu mengidentifikasi seluruh data negatif secara optimal.

Sementara itu, pada kelas positif, model tidak mampu melakukan klasifikasi dengan baik, yang ditunjukkan dengan nilai precision, recall, dan f1-score sebesar 0.00. Hal ini disebabkan oleh jumlah data pada kelas positif yang sangat sedikit dibandingkan dengan kelas lainnya, sehingga model mengalami kesulitan dalam mempelajari pola dari kelas tersebut. Ketidakseimbangan jumlah data (data imbalance) menjadi salah satu faktor utama yang memengaruhi kinerja model, di mana sebagian besar data didominasi oleh kelas netral. Hal ini menyebabkan model cenderung lebih sering mengklasifikasikan data ke dalam kelas netral.

Berikut gambar grafik pie yang ditampilkan menggambarkan distribusi sentimen komentar berdasarkan hasil klasifikasi, di mana kategori “Netral” mendominasi dengan persentase sebesar 76.2%, diikuti oleh “Negatif” sebesar 20.2%, dan “Positif” sebesar 3.6%. Visualisasi ini menunjukkan bahwa sebagian besar masyarakat memberikan tanggapan yang bersifat netral terhadap isu yang dianalisis, sementara sebagian lainnya menunjukkan sentimen negatif, dan hanya sebagian kecil yang memberikan respon positif.

Gambar 5. Diagram Distribusi Kenyamanan



Hasil analisis sentimen masyarakat terhadap isu ijazah palsu yang dikaitkan dengan mantan Presiden Joko Widodo pada platform X (Twitter) divisualisasikan menggunakan word cloud untuk menggambarkan distribusi kata yang paling dominan dalam data tweet, sebagaimana ditunjukkan pada gambar berikut.

- Agustina, D., & Pratama, Y. (2022). Implementasi algoritma Naïve Bayes untuk klasifikasi sentimen opini masyarakat di media sosial. *Jurnal Ilmu Komputer*, 10(2), 210–220.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*. O'Reilly Media.
- Hidayati, N., & Kusuma, R. (2020). Analisis sentimen terhadap isu publik di Twitter menggunakan metode klasifikasi teks. *Jurnal Sistem Informasi*, 16(1), 45–53.
- Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216–225.
- Jurafsky, D., & Martin, J. H. (2021). *Speech and language processing* (3rd ed., draft). Kurniawan, R., & Sari, R. (2022). Analisis sentimen media sosial Twitter menggunakan metode Naïve Bayes. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 9(2), 345–352.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.
- Pratama, A., & Wibowo, A. (2023). Sentiment analysis on Twitter data using NaïveBayes algorithm. *Journal of Information Systems Engineering and Business Intelligence*, 9(1), 45–53.
- Putri, A. N., & Hidayat, T. (2020). Analisis sentimen pengguna Twitter terhadap layanan digital menggunakan algoritma Naïve Bayes. *Jurnal Informatika*, 14(2), 101–110.
- Rahman, F., & Santoso, H. (2022). Sentiment analysis of Twitter data using machinelearning approaches. *International Journal of Advanced Computer Science and Applications*, 13(5), 123–130.
- Saputra, D. A., & Nugroho, Y. (2021). Analisis sentimen opini masyarakat terhadap kebijakan pemerintah menggunakan metode Naïve Bayes. *Jurnal RESTI*, 5(4), 789–796.
- Sari, D. P., & Nugroho, A. (2021). Penerapan TF-IDF dan Naïve Bayes untuk analisis sentimen ulasan produk. *Jurnal RESTI*, 5(3), 567–574.
- Suryadi, R., & Prakoso, B. (2021). Text mining for sentiment analysis on Indonesian social media using Naïve Bayes. *Procedia Computer Science*, 179, 56–63.
- Wijaya, A., & Gunawan, D. (2023). Comparison of Naïve Bayes and SVM for sentiment analysis on social media data. *Journal of Data Science and Its Applications*, 6(1), 1–10.
- Zhang, L., Wang, S., & Liu, B. (2018). *Deep learning for sentiment analysis: A survey*.