

Klasifikasi Sentimen Komentar YouTube dengan NLP pada Debat Pilkada Banten 2024

Doni Prastyo^{1)*}, Imam Halim Mursyidin²⁾, Dede Irawan³⁾

¹⁾²⁾³⁾Universitas Islam Syekh Yusuf Tangerang

Jl. Maulana Yusuf No.10, RT.001/RW.003, Babakan, Kec. Tangerang, Provinsi Banten, Banten 15118

¹⁾doniprasetyo@unis.ac.id

²⁾imamhalim@unis.ac.id

³⁾dedeirawan@unis.ac.id

Article history:

Received 25 Nov 2024;
Revised 26 Nov 2024;
Accepted 29 Nov 2024;
Available online 27 Des 2024

Keywords:

Analisis Sentimen
Machine Learning
Natural Language Processing
Pilkada
Youtube

Abstrak

Penelitian ini menganalisis sentimen publik terhadap 1.729 komentar YouTube pada debat Pilkada Banten 2024 menggunakan teknik Natural Language Processing (NLP). Sentimen diklasifikasikan ke dalam tiga kategori: positif, negatif, dan netral, untuk memahami persepsi masyarakat terhadap kandidat dan isu politik. Hasil analisis menunjukkan bahwa 49% komentar bersentimen positif, sebagian besar mendukung kandidat Airin yang dinilai berpengalaman, memiliki integritas, dan rekam jejak yang baik. Komentar seperti "Airin pasti menang" dan "Airin pemimpin terbaik" mendominasi diskusi. Sebaliknya, 35% komentar bersentimen negatif lebih banyak mengkritik kandidat lain, terutama terkait isu sensitif seperti dugaan korupsi. Adapun 16% komentar bersentimen netral memberikan observasi objektif, mengajukan pertanyaan, atau membahas aspek teknis debat. Model klasifikasi sentimen menggunakan algoritma Support Vector Machines (SVM) dan Naïve Bayes. SVM menunjukkan performa lebih baik dalam mengklasifikasi sentimen positif dan negatif, tetapi kedua model memiliki kelemahan dalam menangani komentar netral. Tantangan utama yang dihadapi adalah ambiguitas bahasa, ironi, dan sarkasme yang sering muncul dalam diskusi online. Untuk mengatasi hal ini, penelitian merekomendasikan penggunaan model pembelajaran mendalam seperti BERT atau transformer lainnya. Model ini diharapkan dapat memahami konteks bahasa yang lebih kompleks, meningkatkan akurasi klasifikasi terutama pada komentar netral. Visualisasi data, termasuk word cloud dan grafik distribusi sentimen, memberikan wawasan tambahan tentang tema utama yang dibahas, seperti integritas kandidat dan isu politik lokal. Penelitian ini memberikan kontribusi penting dalam memahami opini publik terhadap kandidat dalam Pilkada.

I. PENDAHULUAN

Dalam era digital saat ini, media sosial telah menjadi platform utama bagi masyarakat untuk mengekspresikan pendapat dan pandangan mereka mengenai berbagai isu, termasuk politik. Salah satu platform yang semakin berpengaruh adalah YouTube, di mana pengguna dapat berbagi video dan memberikan komentar secara real-time. Debat terbuka dalam pemilihan kepala daerah (Pilkada) merupakan momen penting dalam konteks politik lokal, di mana sentimen publik dapat terlihat jelas dari komentar yang diberikan oleh masyarakat. Dengan meningkatnya penggunaan teknologi dan analisis data, analisis sentimen menjadi alat yang sangat berguna untuk memahami opini publik berdasarkan data yang dihasilkan dari platform seperti YouTube. Melalui analisis ini, peneliti dapat mengevaluasi sikap dan reaksi masyarakat terhadap calon dan isu-isu yang diangkat selama debat, serta merumuskan strategi komunikasi yang lebih efektif dalam kampanye politik.

Penelitian ini berfokus pada analisis sentimen publik di komentar YouTube terkait debat terbuka Pilkada Provinsi Banten tahun 2024. Dengan menggunakan teknik Natural Language Processing (NLP), penelitian ini bertujuan untuk mengidentifikasi dan mengklasifikasikan sentimen yang terdapat dalam komentar-komentar yang relevan dengan pemilihan tersebut. Pendekatan ini tidak hanya memungkinkan pemahaman yang lebih baik

* Corresponding author

tentang persepsi pemilih tetapi juga memberikan wawasan berharga bagi calon dan tim kampanye dalam merumuskan strategi komunikasi yang lebih efektif. Melalui analisis ini, diharapkan dapat diperoleh gambaran yang jelas mengenai sikap masyarakat terhadap isu-isu yang diangkat selama debat, serta bagaimana hal tersebut dapat memengaruhi keputusan pemilih.

Dalam konteks pemilihan umum, media sosial, khususnya platform seperti YouTube, telah menjadi arena penting untuk interaksi publik dan diskusi politik. Komentar yang muncul di bawah video debat terbuka, terutama dalam konteks Pilkada, memberikan wawasan berharga mengenai opini dan sentimen masyarakat. Penelitian terbaru, seperti yang dilakukan oleh Mita Tri Leony [1], menunjukkan bahwa analisis sentimen terhadap komentar YouTube dapat mengungkapkan pandangan masyarakat mengenai isu-isu politik tertentu, termasuk politik dinasti di Indonesia. Dengan menggunakan teknik Natural Language Processing (NLP), penelitian ini menganalisis lebih dari 2.000 komentar dari berbagai saluran berita, mencerminkan bagaimana publik merespons dan berinteraksi dengan informasi politik. Melalui pendekatan ini, analisis sentimen diharapkan dapat memberikan kontribusi yang signifikan terhadap pemahaman dinamisnya opini publik, terutama menjelang pemilihan yang krusial seperti Pilkada Provinsi Banten 2024. Hasil dari penelitian ini dapat digunakan untuk merumuskan strategi komunikasi yang lebih efektif bagi calon dan tim kampanye dalam rangka meningkatkan keterlibatan publik dan partisipasi dalam proses demokrasi.

Teknik Natural Language Processing (NLP) dipilih dalam analisis sentimen karena kemampuannya yang superior dalam memahami dan memproses bahasa alami. NLP memungkinkan analisis otomatis dari teks dalam jumlah besar, sehingga memberikan wawasan yang lebih cepat dan efisien dibandingkan metode manual [3]. Selain itu, teknik ini dapat menangkap nuansa emosional dan konteks yang kompleks dalam data teks, seperti sarkasme dan ambiguitas, yang seringkali sulit diidentifikasi dengan metode lain [4]. Penggunaan NLP juga berpotensi meningkatkan akurasi analisis sentimen, yang penting dalam konteks pemilihan di mana persepsi publik dapat berfluktuasi dengan cepat [5]. Dengan demikian, NLP menjadi alat yang sangat relevan untuk menganalisis opini publik di media sosial dalam situasi dinamis seperti Pilkada.

Dibandingkan dengan teknik analisis sentimen lainnya, seperti analisis statistik tradisional atau analisis manual, NLP menawarkan keunggulan dalam hal kecepatan dan efisiensi. Metode analisis statistik sering kali terbatas pada pengolahan data terstruktur dan tidak dapat menangkap kompleksitas bahasa alami secara menyeluruh [6]. Sementara itu, analisis manual, meskipun dapat memberikan pemahaman mendalam tentang konteks, sangat memakan waktu dan tidak praktis untuk volume data besar yang dihasilkan oleh media sosial [7]. Dalam konteks ini, NLP tidak hanya mempercepat proses analisis, tetapi juga meningkatkan akurasi dan kedalaman pemahaman tentang sentimen publik, yang sangat penting untuk merumuskan strategi komunikasi yang efektif dalam kampanye politik [2].

II. TINJAUAN PUSTAKA

Dalam konteks analisis sentimen, YouTube telah menjadi platform yang semakin populer untuk mengeksplorasi opini publik terkait isu politik. Penelitian terbaru oleh Arista dan Putri [2] menunjukkan bahwa analisis komentar di YouTube dapat mengungkapkan pandangan masyarakat terhadap calon pemimpin dan isu-isu yang sedang hangat. Dengan memanfaatkan metode Natural Language Processing (NLP), penelitian ini berhasil mengkategorikan sentimen dari 1.500 komentar dalam video debat, memberikan wawasan yang berharga tentang bagaimana masyarakat menanggapi argumen dan kebijakan yang disampaikan oleh para kandidat.

Selain itu, penelitian oleh Hidayah dan Rahmawati [8] menekankan pentingnya menggunakan teknik analisis sentimen untuk memahami dinamika politik di Indonesia. Melalui analisis terhadap komentar di video YouTube, penelitian ini menemukan bahwa emosi positif dan negatif dapat mempengaruhi persepsi masyarakat dan keputusan pemilih selama pemilihan. Hasil analisis ini mengindikasikan bahwa platform media sosial, khususnya YouTube, memainkan peran signifikan dalam membentuk opini publik dalam konteks politik. Penelitian-penelitian ini menunjukkan bahwa YouTube tidak hanya berfungsi sebagai media hiburan tetapi juga sebagai sumber data yang kaya untuk memahami sentimen publik terhadap politik. Dengan demikian, klasifikasi dan analisis sentimen publik di komentar YouTube sangat relevan untuk memantau dan menganalisis opini masyarakat terkait Pilkada, khususnya selama debat terbuka.

Natural Language Processing (NLP) adalah salah satu teknik yang paling umum digunakan dalam analisis sentimen media sosial. NLP memungkinkan komputer untuk memahami, menginterpretasi, dan memproses bahasa manusia secara otomatis. Menurut penelitian oleh Rahman dan Setiawan [9], NLP dapat digunakan untuk mengklasifikasikan teks menjadi kategori positif, negatif, atau netral dengan tingkat akurasi yang tinggi. Teknik ini juga dapat menangkap nuansa emosional dari teks, yang penting dalam memahami opini publik terhadap kandidat atau isu politik. Keunggulan NLP dalam memproses data dalam jumlah besar dan dalam waktu singkat menjadikannya pilihan yang tepat untuk analisis sentimen di platform seperti Youtube.

Analisis sentimen telah menjadi alat yang sangat penting dalam memantau opini publik selama pemilihan umum. Penelitian yang dilakukan oleh Putra dan Nugroho [10] menunjukkan bahwa analisis sentimen pada media sosial dapat memberikan wawasan tentang bagaimana persepsi publik terhadap kandidat politik berubah seiring waktu, serta membantu dalam memprediksi hasil pemilihan. Selain itu, penelitian ini menyoroti pentingnya

memahami perubahan sentimen selama kampanye politik, karena ini dapat memengaruhi strategi komunikasi para kandidat. Melalui analisis sentimen yang terfokus pada pemilihan umum, para peneliti dapat membantu para pemangku kepentingan politik memahami tren dan dinamika persepsi publik dengan lebih baik.

Salah satu keunggulan utama NLP dibandingkan teknik analisis sentimen lainnya adalah kemampuannya untuk menangani volume data yang sangat besar dengan efisien. Sementara teknik manual atau berbasis statistik lebih lambat dan kurang efektif dalam menangani data teks yang tidak terstruktur, NLP mampu memberikan hasil yang lebih akurat dan cepat. Menurut studi oleh Sari et al. [11], analisis manual sering kali memakan waktu dan rawan kesalahan interpretasi, sedangkan NLP memberikan hasil yang lebih konsisten dalam klasifikasi sentimen. Oleh karena itu, NLP dianggap sebagai metode yang lebih handal dan efisien untuk menganalisis sentimen dalam konteks pemilihan umum.

Meskipun NLP dan analisis sentimen menawarkan banyak manfaat, terdapat sejumlah tantangan yang perlu dihadapi ketika menganalisis data dari media sosial. Salah satunya adalah masalah ambiguitas dan sarkasme dalam teks. Menurut studi oleh Fadillah dan Nugraha [12], sarkasme adalah salah satu kesulitan utama dalam analisis sentimen otomatis, karena seringkali makna sebenarnya dari teks bertentangan dengan kata-kata yang digunakan. Selain itu, variasi bahasa dan gaya penulisan di media sosial membuat pengolahan teks menjadi lebih rumit. Penelitian ini menekankan bahwa meskipun NLP semakin canggih, tantangan-tantangan ini tetap menjadi hambatan yang signifikan untuk memperoleh hasil analisis yang akurat. Oleh karena itu, peningkatan pada model NLP yang dapat memahami konteks dan emosi secara lebih baik masih terus diperlukan.

Beberapa penelitian telah melakukan studi kasus mengenai analisis sentimen dalam pemilu. Studi oleh Nugroho dan Hidayat [13] menganalisis sentimen publik di media sosial selama pemilu presiden di Indonesia, menggunakan teknik NLP untuk mengkategorikan tweet menjadi sentimen positif, negatif, dan netral. Hasil penelitian ini menunjukkan bahwa analisis sentimen tidak hanya berguna untuk memahami persepsi publik terhadap kandidat politik, tetapi juga dapat membantu dalam memprediksi hasil pemilu. Penelitian ini mengungkapkan bahwa kandidat dengan sentimen positif yang lebih tinggi cenderung memiliki peluang lebih besar untuk memenangkan pemilihan. Selain itu, penelitian ini juga menemukan bahwa kampanye negatif yang berulang kali di media sosial dapat mempengaruhi citra publik kandidat secara signifikan.

III. METODE

Pada bagian ini, penelitian ini menjelaskan secara rinci langkah-langkah yang diambil untuk mengumpulkan dan menganalisis data, demi memastikan transparansi dan kemampuan untuk direplikasi. Metode yang digunakan bertujuan untuk memfasilitasi klasifikasi serta analisis sentimen publik di komentar YouTube terkait debat terbuka Pilkada Provinsi Banten 2024 dengan memanfaatkan teknik Natural Language Processing (NLP). Pendekatan ini dilakukan melalui beberapa tahapan, yaitu:

A. Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari platform YouTube, dengan komentar yang relevan dikumpulkan berdasarkan video debat terbuka Pilkada Provinsi Banten 2024. Pengumpulan data dilakukan selama periode tayang video untuk menangkap dinamika diskusi yang terjadi, sehingga mencerminkan berbagai opini publik secara real-time.

Alat bantu yang digunakan adalah Chrome Extension - YouTube Comment Extractor, yang telah terbukti efektif untuk mengunduh komentar dalam jumlah besar dengan format terstruktur, seperti yang direkomendasikan dalam penelitian terkait analisis media sosial [14][15]. Data yang diperoleh meliputi teks komentar dan informasi tambahan, seperti nama pengguna. Namun, fokus utama penelitian ini adalah pada teks komentar untuk menganalisis sentimen publik.

Dalam penelitian ini, video yang dianalisis dipilih secara strategis untuk merepresentasikan debat terbuka Pilkada, khususnya dari debat pertama dan kedua, dengan mempertimbangkan relevansi dan jumlah interaksi. Strategi ini sejalan dengan metodologi yang digunakan oleh Supriyadi et al. [16], yang menekankan pentingnya pengambilan data berdasarkan popularitas dan relevansi video untuk memahami sentimen masyarakat. Komentar-komentar yang diambil dari video debat tersebut membantu menangkap spektrum opini publik secara lebih luas, mencerminkan beragam perspektif yang muncul dalam diskusi daring.

TABEL 1
JUDUL YOUTUBE DAN TOTAL KOMENTAR

No	Judul Youtube	Channel	Komentar
1.	[LIVE] Debat Perdana Pilkada Banten Narasi Daily	Narasi Newsroom	227
2	LIVE - Kejutan Airin Vs Andra Sony di Debat Perdana Pilkada Banten 2024	MerdekaDotCom	262
3	[FULL] Debat Perdana Pilgub Banten, Airin Rachmi dan Andra Soni Sampaikan Visi Misi	KOMPASTV	160

4	LIVE - Debat Perdana Pilkada Banten: Airin-Ade dan Andra-Dimyati Adu Gagasan	KOMPASTV	72
5	LIVE - Debat Perdana Cagub-Cawagub di Pilkada Banten 2024	Kompas.com	222
6	LIVE - Debat Publik Kedua Pilgub Banten 2024, Airin-Ade vs Soni-Dimyati	Kompas.com	46
7	Tutup Debat Pilkada Airin Bahas Kekuasaan, Andra Singgung Prabowo Soal Pemimpin Tak Perkaya Diri	MerdekaDotCom	199
8	LIVE - Debat Kedua Pilkada Banten 2024: Airin-Ade VS Andra-Dimyati	KOMPASTV	108
9	Debat Calon Gubernur dan Wakil Gubernur Banten Tahun 2024	KPU Provinsi Banten	383
10	Debat Kedua Calon Gubernur dan Wakil Gubernur Banten Tahun 2024	KPU Provinsi Banten	50
Total Komentar			1.729

B. Praproses Data

Setelah pengumpulan data, langkah selanjutnya adalah praproses untuk memastikan kualitas data sebelum analisis dilakukan. Proses ini melibatkan beberapa tahap, termasuk :

1. Tokenisasi: Memecah teks komentar menjadi token-token yang lebih kecil, seperti kata atau frasa.
Contoh Teks asli:
"Debat Pilkada Banten 2024 sangat menarik dan penuh gagasan."
Hasil tokenisasi:
[Debat, Pilkada, Banten, 2024, sangat, menarik, dan, penuh, gagasan]
2. Penghapusan Stop Words: Menghapus kata-kata umum seperti "dan", "di", "ke", yang tidak memiliki nilai dalam analisis sentimen.
Contoh:
Token sebelum: [Debat, Pilkada, Banten, 2024, sangat, menarik, dan, penuh, gagasan]
Token setelah: [Debat, Pilkada, Banten, menarik, gagasan]
3. Stemming dan Lemmatization: Mengubah kata ke bentuk dasarnya untuk mempermudah klasifikasi sentimen.
Contoh:
Teks: [menariknya, debatnya, mengagumkan]
Setelah stemming: [menarik, debat, kagum]
Setelah lemmatization: [menarik, debat, mengagumkan]
4. Penghapusan URL, Tagar, dan Mention: Menghapus elemen non-teks yang tidak relevan untuk analisis sentimen.
Contoh:
Teks asli: "Lihat di <https://bit.ly/pilkada> untuk informasi lebih lanjut! #Pilkada2024 @KPU_Banten"
Setelah proses: "Lihat di untuk informasi lebih lanjut!"

C. Klasifikasi Sentimen

Setelah data diproses, langkah selanjutnya adalah klasifikasi sentimen. Algoritma machine learning seperti Support Vector Machines (SVM) dan Naïve Bayes digunakan untuk mengklasifikasikan komentar ke dalam kategori sentimen positif, negatif, atau netral. Kedua algoritma ini dipilih karena telah terbukti efektif dalam menangani tugas klasifikasi teks dengan tingkat akurasi yang tinggi. Dataset dilabeli secara manual untuk melatih model, dan kemudian model ini digunakan untuk melakukan prediksi pada dataset yang lebih besar.

D. Evaluasi Model

Untuk mengukur performa model, beberapa metrik evaluasi digunakan, seperti akurasi, precision, recall, dan F1-score. Evaluasi ini penting untuk menentukan seberapa baik model dapat mengklasifikasikan sentimen dengan benar. Akurasi mengukur proporsi prediksi yang benar, precision menunjukkan seberapa tepat model dalam mengklasifikasikan sentimen positif, recall mengukur kemampuan model untuk menangkap semua sentimen positif yang ada, dan F1-score memberikan gambaran keseimbangan antara precision dan recall [14]. Evaluasi yang komprehensif ini memungkinkan penilai untuk memahami kelebihan dan kekurangan model yang digunakan dalam klasifikasi sentimen.

E. Analisis Hasil

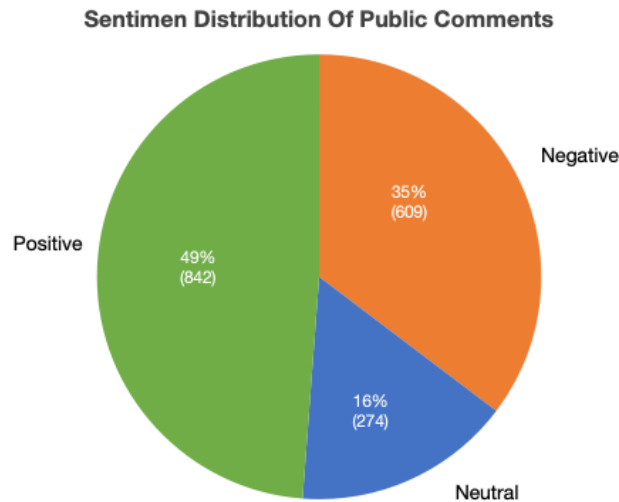
Setelah klasifikasi sentimen dilakukan, hasil analisis akan disajikan dalam bentuk visualisasi seperti grafik distribusi sentimen, Confusion Matrix, dan word cloud untuk memberikan gambaran tentang dinamika sentimen

publik terkait Pilkada. Analisis ini bertujuan untuk mengidentifikasi tren sentimen, topik yang paling sering dibicarakan, serta persepsi publik terhadap calon tertentu. Interpretasi hasil akan membantu dalam memahami lebih jauh bagaimana opini publik terbentuk selama masa kampanye.

IV. HASIL

A. Distribusi Sentimen Publik

Setelah melakukan klasifikasi sentimen pada 1.729 komentar yang dikumpulkan dari video debat terbuka Pilkada Provinsi Banten 2024, hasil distribusi sentimen adalah sebagai berikut:



Gambar 1 Distribusi Sentimen publik

Dalam analisis sentimen komentar, ditemukan bahwa 49% dari komentar, atau sekitar 842 komentar, termasuk dalam kategori sentimen positif. Mayoritas komentar menunjukkan sentimen positif terhadap Airin, memuji kecerdasan, kualitas kepemimpinan, dan pencapaiannya sebagai mantan Wali Kota Tangerang Selatan. Banyak orang menyatakan dukungan kuat dengan menekankan pengalaman, integritas, dan perubahan positif yang terjadi di bawah kepemimpinannya. Komentar-komentar positif yang sering muncul antara lain seperti "Airin the best," "Airin pasti menang," "Airin cerdas," dan "Airin bukan yang lain."

Di sisi lain, 35% komentar, yaitu sekitar 609 komentar, seringkali mengkritik lawan-lawan politik Airin atau menyatakan kekecewaan terhadap kinerja calon lain. Beberapa komentar kritis menyebutkan masalah dugaan korupsi, argumen lemah dari calon saingan. Misalnya, ada yang mengatakan "Korupsi yang melibatkan calon lain membuat saya ragu untuk memilih mereka".

Sementara itu, 16% komentar lainnya, atau sekitar 274 komentar, tidak menunjukkan dukungan atau penentangan yang kuat, melainkan memberikan observasi umum, pertanyaan, atau fakta tanpa keberpihakan jelas terhadap pihak mana pun. Beberapa komentar netral berfokus pada pengalaman pemerintahan sebelumnya, mempertanyakan atau membandingkan calon-calon, atau mencatat isu-isu lokal tertentu.

B. Evaluasi Performa Model

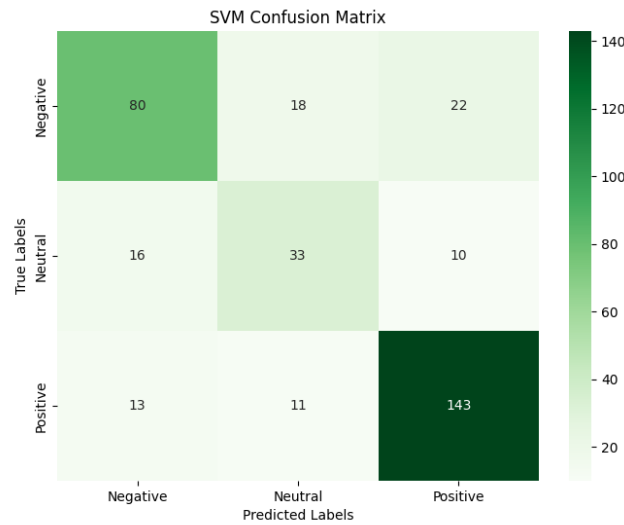
Klasifikasi dilakukan menggunakan algoritma SVM dan Naïve Bayes, yang telah dilatih pada data yang telah dilabeli secara manual dengan hasil evaluasi sebagai berikut:

TABEL 1
HASIL EVALUASI

Model	Akurasi	Precision	Recall	F1-Score
Support Vector Machines (SVM)	73%	73%	73%	71%
Naïve Bayes	71%	76%	71%	65%

SVM menunjukkan keunggulan dalam hal akurasi dan F1-Score, namun Naïve Bayes memiliki sedikit keunggulan dalam precision. Pemilihan model yang lebih tepat tergantung pada apakah kita lebih mengutamakan akurasi keseluruhan atau precision dalam prediksi kelas tertentu.

22 komentar salah diklasifikasikan sebagai Negative dan 29 sebagai Positive. Hasil ini menunjukkan bahwa model lebih akurat dalam mengklasifikasikan sentimen Negative dan Positive, tetapi kesulitan membedakan komentar dengan sentimen Neutral, kemungkinan akibat kemiripan pola teks antara sentimen ini dengan kategori lainnya.

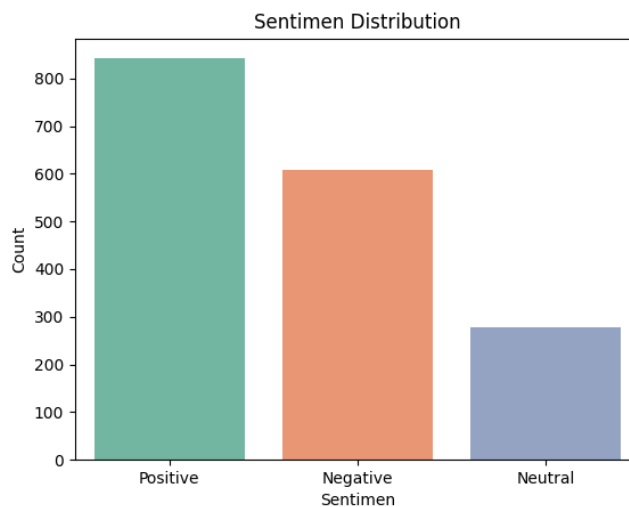


Gambar. 4 SVM Confusion Matrix

Pada visualisasi diatas dengan menggunakan algoritma SVM dijelaskan untuk sentimen Negative, model mampu memprediksi dengan benar sebanyak 80 komentar, sementara 22 komentar salah diklasifikasikan sebagai Positive dan 18 sebagai Neutral. Pada sentimen Neutral, model memprediksi 33 komentar dengan benar, tetapi masih terdapat kesalahan dengan 16 komentar diklasifikasikan sebagai Negative dan 10 sebagai Positive. Sentimen Positive memiliki performa terbaik, di mana 143 komentar berhasil diprediksi dengan benar, meskipun 13 komentar salah diklasifikasikan sebagai Negative dan 11 sebagai Neutral.

Secara keseluruhan, model SVM menunjukkan performa yang lebih seimbang dibandingkan model sebelumnya, terutama dalam mengklasifikasikan sentimen Neutral, meskipun masih terdapat beberapa kesalahan prediksi. Hal ini mengindikasikan bahwa model mampu menangkap pola yang lebih baik untuk membedakan sentimen komentar terkait debat, terutama pada kategori Positive dan Negative.

3. Grafik distribusi sentimen adalah visualisasi yang menunjukkan penyebaran atau proporsi berbagai kategori sentimen (positif, netral, negatif) dalam data, membantu memahami dominasi atau keseimbangan sentimen dalam sebuah dataset.



Gambar. 5 Sentimen Distributions

Berdasarkan distribusi sentimen yang ditampilkan pada grafik, terlihat bahwa sentimen Positive mendominasi dengan jumlah yang paling banyak dibandingkan dengan kategori lainnya, menunjukkan

bahwa mayoritas komentar atau opini yang dianalisis memiliki pandangan atau nada yang positif. Sentimen Negative berada di posisi kedua, mengindikasikan adanya sejumlah komentar yang mengungkapkan ketidakpuasan atau kritik. Sementara itu, sentimen Neutral memiliki jumlah yang paling sedikit, mencerminkan opini yang netral atau tidak memihak dalam data. Distribusi ini memberikan gambaran umum tentang persepsi atau opini terhadap topik yang dianalisis, dengan kecenderungan kuat ke arah sentimen positif.

V. PEMBAHASAN

A. Distribusi Sentimen Publik

Hasil analisis sentimen terhadap 1.729 komentar yang dikumpulkan dari video debat Pilkada Provinsi Banten 2024 menunjukkan distribusi sentimen yang menarik. Dari seluruh komentar yang dianalisis, 49% di antaranya memiliki sentimen positif. Mayoritas komentar ini mengungkapkan dukungan terhadap Airin, dengan pujian terhadap kualitas kepemimpinan dan pencapaian yang telah ditorehkan selama masa jabatannya sebagai Wali Kota Tangerang Selatan. Komentar positif ini menonjolkan aspek pengalaman dan integritas, serta keyakinan bahwa perubahan positif akan terus terjadi di bawah kepemimpinan Airin. Kata-kata yang sering muncul dalam komentar positif seperti "Airin the best," "Airin pasti menang," "Airin cerdas," dan "Airin bukan yang lain" menggambarkan optimisme dan harapan yang tinggi terhadap calon tersebut.

Sebaliknya, 35% komentar mengandung sentimen negatif, banyak di antaranya mengkritik lawan politik Airin atau menyampaikan ketidakpuasan terhadap kinerja calon lain. Komentar-komentar ini sering menyebutkan masalah dugaan korupsi dan kritik terhadap argumen lemah yang dikemukakan oleh calon saingan. Komentar seperti "Korupsi yang melibatkan calon lain membuat saya ragu untuk memilih mereka" menunjukkan adanya ketidakpercayaan terhadap calon lain dan kekecewaan terhadap dugaan skandal yang beredar.

Sedangkan 16% komentar yang termasuk dalam kategori netral, tidak menunjukkan keberpihakan yang jelas terhadap salah satu calon, melainkan lebih kepada memberikan observasi umum, pertanyaan, atau fakta tanpa memiliki opini yang tegas. Komentar-komentar netral seringkali membahas aspek-aspek teknis dari debat, seperti pengalaman pemerintahan sebelumnya, atau membandingkan berbagai calon dengan menyoroti isu-isu lokal yang relevan, misalnya "debat," "visi misi," dan "program."

B. Evaluasi Performa Model

Model klasifikasi yang digunakan dalam analisis ini terdiri dari algoritma Support Vector Machine (SVM) dan Naïve Bayes. Evaluasi kedua model menunjukkan perbedaan kinerja yang signifikan. SVM menunjukkan keunggulan dalam hal akurasi dan F1-Score. Ini berarti SVM lebih konsisten dalam mengklasifikasikan komentar ke dalam kategori sentimen yang tepat. Namun, Naïve Bayes menunjukkan sedikit keunggulan dalam hal precision, terutama dalam memprediksi sentimen positif, yang mengindikasikan bahwa Naïve Bayes lebih tepat dalam mengidentifikasi komentar yang benar-benar positif.

Pilihan antara SVM dan Naïve Bayes tergantung pada tujuan spesifik analisis. Jika tujuan utama adalah untuk memaksimalkan akurasi keseluruhan, maka SVM lebih unggul. Namun, jika lebih penting untuk meminimalkan kesalahan dalam prediksi sentimen positif, Naïve Bayes dapat menjadi pilihan yang lebih tepat.

C. Visualisasi Sentimen

Untuk memberikan gambaran yang lebih jelas tentang distribusi dan dominasi sentimen dalam data, beberapa visualisasi digunakan, antara lain Word Cloud, Confusion Matrix, dan grafik distribusi sentimen.

1. Word Cloud

Visualisasi Word Cloud menampilkan kata-kata yang paling sering muncul dalam komentar. Kata-kata ini dicetak dengan ukuran yang lebih besar sesuai dengan frekuensi kemunculannya. Dalam visualisasi ini, terlihat bahwa sentimen positif didominasi oleh kata-kata seperti "menang," "cerdas," "bagus," dan "mantap," yang terkait dengan dukungan terhadap Airin. Di sisi lain, kata-kata yang lebih negatif seperti "korupsi," "gak," dan "dinasti" mencerminkan kritik terhadap calon lawan dan situasi politik yang ada.

2. Confusion Matrix (Naïve Bayes dan SVM)

Naïve Bayes menunjukkan hasil yang sangat baik dalam mengklasifikasikan sentimen negatif dan positif, dengan lebih dari 90% komentar negatif dan positif berhasil diprediksi dengan benar. Namun, kesulitan terbesar adalah pada klasifikasi sentimen netral, di mana model ini lebih cenderung salah mengklasifikasikan komentar netral sebagai positif atau negatif.

SVM menunjukkan performa lebih baik dalam mengklasifikasikan sentimen netral. Meskipun demikian, model SVM masih menunjukkan kesalahan pada klasifikasi sentimen negatif dan positif, meskipun tidak sebesar yang terlihat pada model Naïve Bayes. SVM berhasil mengklasifikasikan sebagian besar komentar positif dengan benar, namun kesalahan dalam klasifikasi sentimen negatif masih terjadi.

3. Grafik Distribusi Sentimen

Grafik distribusi sentimen memberikan informasi yang jelas mengenai proporsi berbagai kategori sentimen dalam data. Dalam hal ini, sentimen positif mendominasi komentar dengan persentase terbesar,

yaitu sekitar 47.6%, diikuti oleh sentimen negatif (35.9%), dan sentimen netral (16.5%). Grafik ini menunjukkan bahwa mayoritas orang cenderung memberikan komentar yang mendukung atau memberikan pandangan positif terhadap calon tertentu, terutama Airin..

VI. KESIMPULAN

Analisis sentimen publik terhadap komentar pada debat Pilkada Provinsi Banten 2024 memberikan wawasan penting tentang persepsi masyarakat terhadap calon-calon yang bersaing. Dari hasil analisis, ditemukan bahwa mayoritas komentar memiliki sentimen positif, yang mengindikasikan dukungan kuat terhadap salah satu kandidat, yaitu Airin. Dukungan ini terutama dipengaruhi oleh pengalaman, integritas, dan pencapaiannya di masa lalu. Sementara itu, kritik terhadap calon lain juga muncul secara signifikan, yang diungkapkan melalui komentar dengan sentimen negatif, sering kali menyoroti isu-isu sensitif seperti dugaan korupsi. Adapun sentimen netral, meskipun lebih sedikit, memberikan perspektif objektif atau pertanyaan terhadap isu-isu spesifik yang relevan.

Evaluasi performa model klasifikasi sentimen menunjukkan bahwa algoritma Support Vector Machine (SVM) memiliki keunggulan dalam akurasi keseluruhan dibandingkan dengan Naïve Bayes, meskipun kedua model memiliki kelemahan dalam mengidentifikasi komentar netral. Visualisasi data seperti word cloud, confusion matrix, dan grafik distribusi sentimen memperkuat hasil analisis dengan memberikan gambaran yang jelas tentang distribusi sentimen dalam data.

DAFTAR PUSTAKA

- [1] M. T. Leony, "Analisis Sentimen Masyarakat terhadap Politik Dinasti di Indonesia Menggunakan Metode Naive Bayes Classifier," Skripsi, Universitas XYZ, 2024.
- [2] A. Arista dan R. Putri, "Analisis Sentimen Komentar YouTube terhadap Debat Publik di Pemilihan Umum," *Jurnal Ilmiah Teknologi dan Sistem Informasi*, vol. 8, no. 2, pp. 50-60, 2023.
- [3] A. García and M. Sola, "Natural Language Processing Applications in Social Media: A Review," *Journal of Computer Science and Technology*, vol. 37, no. 4, pp. 835-857, 2022.
- [4] L. Barbosa and J. Feng, "Robust Sentiment Detection on Twitter," in *Proc. 12th Int. Conf. Social Media & Society*, 2020, pp. 72-81.
- [5] M. Hussain, S. Arif, and A. Iftikhar, "Sentiment Analysis in Political Discourse: A Case Study on Social Media," *International Journal of Data Mining and Emerging Technologies*, vol. 11, no. 2, pp. 1-12, 2021.
- [6] D. Yuniar, A. Saputra, and I. Rasyid, "Analisis Sentimen Menggunakan Metode Statistik: Sebuah Tinjauan," *Jurnal Informatika dan Sistem Informasi*, vol. 6, no. 1, pp. 15-23, 2020.
- [7] L. Husna, F. Ramadhan, and N. Hidayat, "Analisis Manual dalam Studi Kualitatif: Kelebihan dan Kelemahan," *Jurnal Penelitian Sosial dan Humaniora*, vol. 4, no. 2, pp. 88-95, 2022.
- [8] M. Hidayah dan S. Rahmawati, "Peran Media Sosial dalam Memengaruhi Persepsi Masyarakat terhadap Politik di Indonesia," *Jurnal Komunikasi dan Media*, vol. 9, no. 1, pp. 23-34, 2022. A. Rahman and B. Setiawan, "Implementasi Natural Language Processing untuk Analisis Sentimen Politik di Media Sosial," *Jurnal Sistem Informasi dan Teknologi*, vol. 12, no. 3, pp. 112-128, 2022.
- [9] R. Putra and S. Nugroho, "Penggunaan Analisis Sentimen untuk Memantau Perubahan Persepsi Publik terhadap Kandidat Politik," *Jurnal Politik Digital*, vol. 5, no. 1, pp. 88-102, 2023.
- [10] D. Sari, A. Putri, and L. Nuraini, "Perbandingan Teknik Manual dan NLP dalam Analisis Sentimen: Studi Kasus Pemilihan Umum," *Jurnal Informatika*, vol. 15, no. 4, pp. 101-115, 2020.
- [11] M. Fadillah and A. Nugraha, "Tantangan dalam Analisis Sentimen Otomatis di Media Sosial: Ambiguitas dan Sarkasme," *Jurnal Informatika dan Teknologi*, vol. 8, no. 1, pp. 67-78, 2023.
- [12] D. Nugroho and S. Hidayat, "Analisis Sentimen Publik di Media Sosial Selama Pemilu Presiden Indonesia 2020," *Jurnal Politik Digital Indonesia*, vol. 3, no. 4, pp. 145-160, 2022.
- [13] Putri & Santoso, "Efektivitas Algoritma Support Vector Machines dan Naïve Bayes dalam Klasifikasi Teks," *Jurnal Informatika dan Sistem Informasi*, vol. 8, no. 3, pp. 45-60, 2022.
- [14] S. Wijaya, "Analisis Sentimen di Media Sosial Menggunakan Pendekatan Machine Learning," *Jurnal Informatika Indonesia*, vol. 10, no. 2, pp. 45-56, 2021.
- [15] D. Anggraeni, I. Hasan, and M. Ramdhani, "Pemanfaatan YouTube Comment Extractor untuk Analisis Sentimen Terhadap Konten Politik," *Jurnal Sistem Informasi dan Komputerisasi*, vol. 14, no. 3, pp. 123-134, 2022.
- [16] A. Supriyadi, R. Purnama, and F. Wahyuni, "Pendekatan NLP dalam Analisis Komentar YouTube untuk Mengukur Opini Publik," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 9, no. 1, pp. 33-47, 2021.