

Analisis Sentimen pada Sengketa Pilpres 2024 dengan *Multinomial Naïve Bayes* dan *Chi-Square*

I Gede Widnyana^{a1}, AAIN Eka Karyawati^{a2}

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
¹widnyana.2208561016@student.unud.ac.id
²eka.karyawati@unud.ac.id

Abstract

The Constitutional Court officially rejected all appeals against the 2024 Presidential Election results in Jakarta, drawing public attention. People expressed their opinions through social media platforms, including YouTube. Therefore, this research was conducted to analyze public sentiment regarding this political phenomenon. This study uses primary data on comments from 3 YouTube videos that were crawled with a total of 4103 data. The data underwent labeling and then went through 5 stages of text pre-processing. Subsequently, modeling was performed using the Multinomial Naïve Bayes Algorithm with testing on 10 variations of the Chi-Square threshold, followed by validation with 10-fold cross-validation. The research yielded an optimal model performance, including accuracy, precision, recall, and F1-score, stabilizing at 72% with a Chi-Square feature selection of 70% and 90%. Without feature selection, the model's performance was at 71.5%. Model validation with the highest and average K-Fold Cross-Validation scores were sequentially at 74% and 69%. The built model is sufficiently good at analyzing sentiment and provides insights that 52% of Indonesians are dissatisfied with the Constitutional Court's decision based on YouTube comments.

Keywords: *Pilpres, Analysis Sentiment, Multinomial Naïve Bayes, Chi-Square, K-Fold Cross Validation*

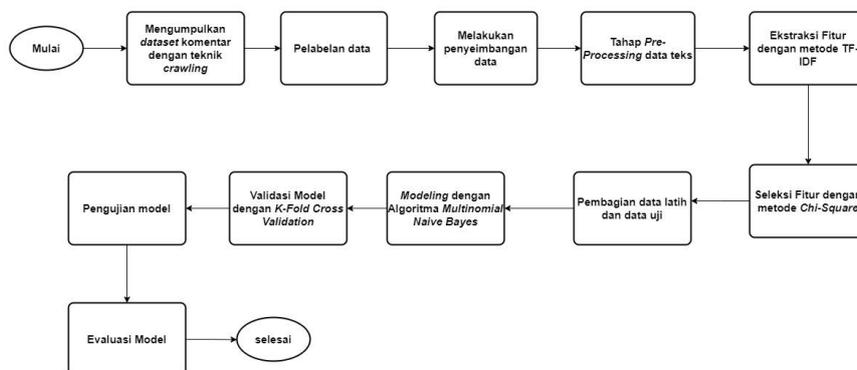
1. Pendahuluan

Pada tanggal 20 Maret 2024, Ketua Komisi Pemilihan Umum (KPU) mengumumkan bahwa pasangan calon presiden-wakil presiden nomor urut 02, Prabowo Subianto-Gibran Rakabuming Raka, dinyatakan sebagai pemenang Pilpres 2024 karena berhasil memperoleh suara terbanyak, yakni sebanyak 96.214.691 suara sah dari total 164.270.475 suara yang masuk [1]. Hasil keputusan tersebut secara resmi mendapatkan gugatan oleh pasangan capres-cawapres Anies Baswedan dan Muhaimin Iskandar serta capres-cawapres Ganjar Pranowo dan Mahmud MD. Namun pada tanggal 22 April 2024, Mahkamah Konstitusi secara resmi menolak seluruh gugatan terhadap hasil Pilpres 2024 melalui sidang pembacaan putusan sengketa hasil Pemilihan Presiden 2024 di Jakarta, sehingga dapat dipastikan bahwa Pilpres 2024 secara sah dimenangkan oleh pasangan capres-cawapres Prabowo Subianto-Gibran Rakabuming Raka [1]. Putusan MK tersebut mengundang perhatian masyarakat sebagai warga negara yang turut andil dalam pelaksanaan demokrasi. Oleh karena itu, masyarakat berhak dalam berpendapat dan menyuarakan opininya terhadap hasil keputusan MK baik secara langsung atau melalui sosial media. Salah satu contohnya adalah masyarakat memberikan tanggapan melalui kolom komentar video putusan sengketa Pilpres 2024 pada *channel* Youtube. Komentar masyarakat ini menghasilkan sebuah sentimen atau emosi dalam bentuk kalimat terhadap fenomena politik ini. Oleh karena itu, penulis perlu melakukan penelitian ini untuk menganalisa sentimen masyarakat terhadap hasil sengketa Pilpres 2024 dengan mengklasifikasikannya menjadi kategori respon positif dan negatif sehingga dapat mengetahui tingkat respon masyarakat terkait isu demokrasi di Indonesia salah satunya adalah pemilihan presiden dan wakil presiden 2024.

Analisis sentimen menjadi salah satu bidang kajian ilmu komputer khususnya dalam pengolahan data teks. Penerapan bidang ilmu ini tercermin pada penelitian terdahulu terkait analisis sentimen masyarakat Indonesia tentang Pilpres 2024 menggunakan Algoritma *Support Vector Machine* dengan *dataset* yang di-*crawling* dari Twitter dan menghasilkan akurasi sebesar 65% yang lebih unggul dalam mengklasifikasikan positif, skor presisi sebesar 69%, *recall* 81% dan *f1-score* sebesar 74% [2]. Selain itu, penelitian kedua dengan metode yang sama dengan penulis namun objek dan dataset yang berbeda diperoleh akurasi 93% dengan menggunakan *Chi-Square* dan 92% tanpa *Chi-Square* [3]. Penelitian ketiga hampir sama dengan penelitian kedua, namun objek dan data yang berbeda serta menghasilkan hasil terbaik dengan akurasi, presisi, *recall*, dan *f-measure* mencapai 86,94%, 80,42%, 80,42%, dan 80,42% berturut-turut pada seleksi fitur 10% [4]. Berdasarkan kajian tersebut, terdapat perbedaan dengan penelitian yang dilakukan penulis yakni terletak pada objek penelitian yang diambil penulis adalah sidang sengketa Pilpres 2024 dengan menggunakan Algoritma *Multinomial Naïve Bayes* dengan seleksi fitur *Chi-Square*. Selain itu terdapat perbedaan *dataset* yang digunakan. Penelitian ini bertujuan untuk menghasilkan tingkat kinerja model melalui evaluasi matriks berdasarkan komentar masyarakat pada platform Youtube Sidang Pilpres 2024 dan mengetahui pengaruh jumlah persentase pada seleksi fitur *Chi-Square* terhadap kinerja model.

2. Metode Penelitian

Pada tahap metode penelitian terdapat gambaran alur tahapan penelitian yang disusun dan dilaksanakan oleh penulis. Pada penelitian ini, *input* yang digunakan berupa komentar dari platform Youtube terkait Sidang Putusan MK Sengketa Pilpres 2024 pada channel Youtube KOMPASTV, CNN Indonesia dan Najwa Shihab. *Input* data komentar tersebut dilabeli dengan 2 kelas label yaitu komentar positif dan komentar negatif untuk mengetahui reaksi masyarakat yang mengarah setuju dan tidak setuju dari putusan sidang MK tersebut. *Output* yang akan diperoleh yakni perbandingan kinerja dari Algoritma *Multinomial Naive Bayes* tanpa seleksi fitur *Chi-Square* dan kinerja dari model menggunakan seleksi fitur *Chi-Square* dengan 10 percobaan penggunaan *Threshold Chi-Square*.



Gambar 1. Rancangan Alur dari Metode Penelitian.

Gambar 1 menjelaskan alur penelitian yang meliputi pengumpulan data primer melalui *crawling*, pelabelan, dan penyeimbangan data dilanjutkan dengan tahap *pre-processing* untuk membersihkan dan normalisasi data, pembobotan fitur menggunakan TF-IDF, dan seleksi fitur menggunakan *Chi-Square* dengan 10 variasi nilai *Threshold* yang berbeda. Kemudian, dilakukan *modeling* menggunakan *Multinomial Naive Bayes*, validasi dengan *K-Fold Cross Validation*, dan evaluasi model untuk menghasilkan matriks evaluasi termasuk akurasi, presisi, *recall*, dan *F-1 score*. Alur kedua memiliki tahap yang sama, namun tidak menerapkan seleksi fitur *Chi-Square*, menggunakan *threshold Chi-Square* sebesar 100% pada tahap seleksi fitur.

2.1 Pengumpulan Data

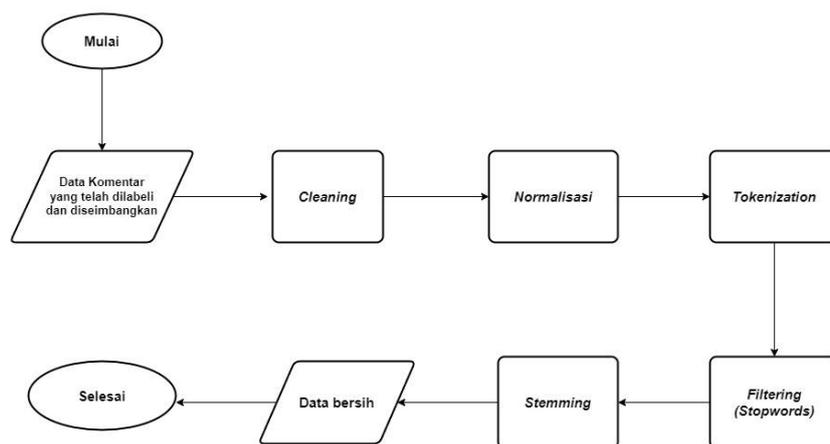
Dataset yang digunakan dalam penelitian ini diperoleh berasal dari komentar video Sidang MK Putusan Pilpres 2024 dari 3 *Channel* Youtube yakni KOMPASTV, CNN Indonesia dan Najwa Shihab menggunakan metode *crawling*. Data yang digunakan adalah komentar masyarakat berbahasa Indonesia dari 3 *Channel* Youtube tersebut. Proses penambangan data komentar dilakukan pada tanggal 27 April 2024 dengan selisih 5 hari dari video yang dipublikasi. Sehingga, data yang diperoleh merupakan data primer dengan periode 22 April sampai 27 April 2024 dan hasil *crawling* data disimpan dan digabungkan dalam bentuk file excel (.xlsx). Data komentar yang diperoleh berjumlah 4103 data yang merupakan gabungan dari komentar video pada 3 *channel* Youtube yang berbeda.

2.2 Pelabelan dan Penyeimbangan Data

Dataset yang telah di-*crawl* kemudian diberi label dengan dua kelas: label 1 untuk positif dan label 0 untuk negatif untuk menggambarkan respon masyarakat terhadap keputusan MK terkait Pilpres 2024. Pemberian label pada data dilakukan secara manual. Hasilnya yakni terdapat 1536 data positif dan 2567 data negatif serta perlu dilakukan penyeimbangan karena ketidakseimbangan antara distribusi data antar kedua kelas tersebut. Teknik penyeimbangan digunakan untuk mengurangi data di kelas mayoritas tanpa menghilangkan informasi penting. Setelah penyeimbangan, kedua kelas memiliki 1536 data, sehingga total *dataset* adalah 3102 data. Penyeimbangan data dilakukan untuk meningkatkan kualitas *model machine learning*, mengurangi bias, dan meningkatkan interpretabilitas data.

2.3 Text Pre-processing

Text pre-processing adalah langkah pertama dalam pemrosesan teks yang bertujuan mempersiapkan data agar siap untuk tahap-tahap berikutnya [5]. Proses ini melibatkan teknik untuk membersihkan, memodifikasi, dan mengorganisir teks agar dapat diolah oleh algoritma NLP (*Natural Language Processing*) atau model pembelajaran mesin [5]. Tujuannya adalah meningkatkan kualitas data dengan menghilangkan informasi tidak relevan, mengatasi variasi dalam representasi teks, dan memastikan konsistensi dalam format dan struktur data. Tahapannya mencakup berbagai teknik seperti yang ditunjukkan dalam gambar 2.



Gambar 2. Rangkaian Teknik *Preprocessing*.

Pada penelitian ini, tahap *text pre-processing* secara detail dijabarkan sebagai berikut.

a. *Cleaning*

Cleaning adalah tahap awal dalam pemrosesan teks yang bertujuan untuk membersihkan dokumen teks dari berbagai jenis *noise* atau gangguan [5]. Ini mencakup penghapusan elemen yang tidak relevan seperti duplikasi kata, angka, karakter selain huruf alfabet, serta

tanda baca seperti titik, koma, tanda tanya, dan tanda seru. Selain itu, dalam tahap cleaning ini juga dilakukan *case folding*, yang mana seluruh huruf dalam dokumen teks diubah menjadi huruf kecil (*lowercase*). Tujuan dari tahap *cleaning* adalah untuk menyederhanakan teks dan menghilangkan elemen yang tidak diperlukan agar dokumen teks dapat diproses lebih efisien dalam tahap analisis selanjutnya.

b. *Normalization*

Normalization adalah tahap dalam pemrosesan teks yang bertujuan untuk mengembalikan kata-kata dalam bahasa yang tidak baku atau tidak standar ke bentuk yang sesuai dengan tata bahasa yang dianggap baku, seperti yang terdapat dalam Kamus Besar Bahasa Indonesia (KBBI) [5]. Tujuan utamanya adalah untuk memperoleh representasi yang konsisten dan baku dari kata-kata dalam teks, sehingga memudahkan analisis dan pengolahan lebih lanjut. Proses normalisasi dapat melibatkan penggantian kata-kata slang, singkatan, atau kata-kata yang tidak formal dengan bentuk yang lebih formal dan baku.

c. *Filtering*

Filtering adalah tahap dalam pemrosesan teks yang bertujuan untuk menghapus kata-kata yang dianggap umum atau tidak informatif dalam konteks tertentu seperti kata-kata penghubung, kata-kata waktu, dan kata-kata lainnya yang mungkin tidak memberikan kontribusi signifikan terhadap analisis teks yang dilakukan [5]. Tujuannya adalah untuk menyederhanakan teks dan fokus pada informasi yang lebih relevan atau penting.

d. *Tokenization*

Tokenization adalah proses dalam pemrosesan teks yang bertujuan untuk membagi atau memisahkan sebuah teks menjadi unit-unit diskret yang lebih kecil, yang disebut token [5]. Token dapat berupa kata-kata, frasa, atau karakter individual, tergantung pada tingkat granularitas yang diinginkan. Tujuan dari *tokenization* adalah untuk menyederhanakan teks dan mengubahnya menjadi representasi yang lebih terstruktur sehingga dapat diolah lebih lanjut oleh komputer.

e. *Stemming*

Stemming adalah tahap dalam pemrosesan teks yang bertujuan untuk menghasilkan kata dasar (*stem*) dari suatu token dengan cara menghilangkan awalan, akhiran, sisipan, dan kombinasi dari awalan dan akhiran [5]. Tujuannya adalah untuk mengubah variasi kata yang berbeda ke dalam bentuk dasarnya, sehingga menciptakan kata-kata yang memiliki akar yang sama untuk diidentifikasi sebagai bentuk yang sama.

2.4 Ekstraksi Fitur *Term Frequency Inverse Document Frequency (TF-IDF)*

TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah metode pembobotan kata yang digunakan dalam pemrosesan teks untuk mengevaluasi pentingnya sebuah kata dalam sebuah dokumen [6]. Metode ini menggabungkan dua konsep utama: *term frequency* (TF) dan *inverse document frequency* (IDF) [6]. Ekstraksi fitur atau pemberian bobot pada setiap kata bertujuan untuk menilai seberapa sering sebuah kata muncul dalam dokumen. Hal ini dilakukan dengan memberikan nilai yang menggambarkan frekuensi kemunculan kata tersebut dalam teks. Pembobotan kata dengan TF-IDF memiliki tahapan-tahapan dengan rumus matematik sebagai berikut:

- a. Rumus matematis untuk menghitung *term frequency* (tf).

$$tf(t, d) = \left(\frac{f(t, d)}{n} \right) \quad (1)$$

- b. Rumus matematis untuk menghitung IDF

$$idf(t) = \log \left(\frac{n}{1 + df(t)} \right) \quad (2)$$

- c. Rumus matematis untuk menghitung bobot TF-IDF

$$w_{t,d} = tf_{t,d} \times idf_t \quad (3)$$

2.5 Seleksi Fitur *Chi-Square*

Proses seleksi fitur bertujuan untuk mengurangi jumlah *noise* dan menghilangkan fitur yang dianggap kurang relevan dalam suatu sistem [4]. *Chi-Square* dipilih karena dapat membantu menilai tingkat independensi suatu fitur terhadap kelas label, sehingga menjadi teknik berguna dan dapat diandalkan dalam diskritisasi dan pemilihan fitur, terutama untuk atribut numerik [3]. Fitur dengan nilai *chi-square* rendah dianggap memiliki dampak yang minim terhadap kinerja sistem, sehingga penggunaannya dapat mengurangi beban sistem [3]. Perhitungan *Chi-Square* disajikan pada persamaan 4 berikut.

$$X^2(t, c) = \frac{N(AD - CB)^2}{(A + C)(B + D)(A + B)(C + D)} \quad (4)$$

Keterangan:

- N = total dokumen secara keseluruhan,
- A = total dokumen yang mempunyai term t pada kelas c ,
- B = total dokumen yang mempunyai term t namun tidak terdapat pada kelas c ,
- C = total dokumen yang tidak mempunyai term t pada kelas c ,
- D = total dokumen yang tidak mempunyai term t tetapi tidak terdapat pada kelas c .

2.6 Pembagian Data

Sebelum menerapkan Algoritma *Multinomial Naïve Bayes*, data dibagi untuk modeling setelah tahap seleksi dan ekstraksi fitur selesai. *Dataset* awal berjumlah 3102 data setelah penyeimbangan, kemudian dibersihkan dengan menghapus 29 data kosong, sehingga tersisa 3703 data. Selanjutnya, data dibagi menjadi 80% data latih dan 20% data uji. Data latih terdiri dari 1565 data (1250 negatif, 1209 positif), sedangkan data uji terdiri dari 615 data (315 negatif, 300 positif). Setelah itu, dilakukan klasifikasi menggunakan Algoritma *Multinomial Naïve Bayes*.

2.7 *Multinomial Naïve Bayes*

Multinomial Naïve Bayes adalah bagian dari *Naïve Bayes* yang termasuk dalam kategori *supervised learning* [7]. Algoritma ini menggunakan prinsip distribusi multinomial dan biasanya digunakan dalam kasus analisis teks dengan mengubah teks ke dalam bentuk nominal yang dapat dihitung menggunakan nilai bilangan bulat [7]. Algoritma ini menghitung probabilitas sebuah dokumen terhadap kelas tertentu, sebagaimana ditunjukkan dalam suatu persamaan 5 berikut.

$$P(C) = \frac{N_C}{N} \quad (5)$$

Berdasarkan persamaan (5) maka N_C sebagai jumlah kelas C untuk keseluruhan dokumen dan N merupakan jumlah seluruh dokumen [7]. Formula Probabilitas dari kata ke- n ditentukan dengan menggunakan persamaan 6 berikut.

$$P(X_n | C) = \frac{N_{x_n, c} + \alpha}{N(C) + V} \quad (6)$$

Dalam konteks ini, $N_{x_n, c}$ mewakili jumlah kemunculan *term* X_n dalam seluruh data latih untuk kelas C , sedangkan $N(C)$ adalah total *term* dalam seluruh data latih untuk kelas C . Parameter *smoothing Laplace* α digunakan untuk menangani kata-kata yang tidak muncul dalam data latih. Selain itu, V merupakan jumlah total kata dalam data latih [7]. Rumus Multinomial yang digunakan dalam pembobotan TF-IDF adalah sebagai berikut.

$$P(X_n | C) = \frac{\sum_{d \in C} tf(X_n, d) + \alpha}{\sum_{d \in C} N_{d \in C} + V} \quad (7)$$

Dalam konteks ini, $\sum_{d \in C} tf(X_n, d)$ mengacu pada jumlah pembobotan *term* X_n dari seluruh data latih dalam kelas C , sedangkan $\sum_{d \in C} N_d$ mengacu pada jumlah bobot dari seluruh *term* dalam data latih dalam kelas C [7].

2.8 Pengujian

Proses pengujian dilakukan untuk mengevaluasi kinerja model dalam menganalisis sentimen komentar. Model diuji dengan mengubah nilai *threshold Chi-Square* untuk mengevaluasi pengaruh seleksi fitur terhadap kinerja. Seleksi fitur dilakukan dengan variasi rasio penggunaan fitur dari 10% hingga 100% dengan kelipatan 10. Hasil klasifikasi dibandingkan dengan label sebenarnya untuk menghitung nilai (TP) *True Positive*, (TN) *True Negative*, (FP) *False Positive*, dan (FN) *False Negative*, yang digunakan untuk mengevaluasi kinerja model secara keseluruhan.

2.9 K-Fold Cross Validation

Cross-validation adalah metode dalam *data mining* untuk mendapatkan nilai akurasi optimal dengan membagi data menjadi data uji dan data latih [8]. Salah satu jenis yang umum adalah *K-Fold Cross Validation*, di mana *dataset* dibagi menjadi K partisi acak. Proses evaluasi dilakukan K kali di mana setiap kali, satu partisi digunakan sebagai data uji dan yang lainnya sebagai data latih [8]. Hal itu berkaitan dengan penggunaan seluruh *dataset* untuk evaluasi dan memberikan estimasi akurasi yang lebih baik untuk model yang dibangun.

2.10 Evaluasi

Evaluasi dihasilkan setelah melakukan pengujian menggunakan *Confusion Matrix* (CM). *Confusion Matrix* adalah salah satu metode yang digunakan untuk mengukur kinerja suatu metode klasifikasi dengan membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang sebenarnya [7].

Tabel 1. Gambaran Tabel Kinerja Model dalam *Confusion Matrix*.

Kelas	Klasifikasi Positif	Klasifikasi Negatif
Positif	TP	FN
Negatif	FP	TN

Keterangan:

- *True Positive* (TP) = jumlah data yang positif yang terprediksi benar (positif).
- *True Negative* (TN) = jumlah data yang negatif yang terprediksi benar (negatif).
- *False Positive* (FP) = jumlah data yang positif namun terprediksi salah (negatif).
- *False Negative* (FN) = jumlah data yang berlabel negatif tetapi terprediksi salah (positif).

Evaluasi dilakukan menggunakan data *testing* yang telah dibagi sebelumnya. Pada tahap evaluasi, performa dari setiap model akan dihitung untuk mencari model dengan performa tertinggi, yang diukur berdasarkan akurasi, presisi, *recall*, dan *F1-score* [7]. Akurasi dihitung menggunakan *confusion matrix*, yang merupakan matriks dua dimensi yang menunjukkan hasil prediksi dari model dan hasil sebenarnya. Untuk menghitung performa model melalui akurasi, presisi, *recall*, dan *F1-score*, digunakan rumus-rumus berikut ini.

$$\text{Nilai Accuracy} = \frac{TP + TN}{(TP + FP + TN + FN)} \quad (8)$$

$$\text{Nilai Precision} = \frac{TP}{(TP + FP)} \quad (9)$$

$$\text{Nilai Recall} = \frac{TP}{(TP + FN)} \quad (10)$$

$$\text{Nilai F1-Score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (11)$$

3. Hasil dan Diskusi

Penelitian ini bertujuan untuk mengklasifikasikan komentar masyarakat menjadi sentimen positif atau negatif menggunakan Algoritma *Multinomial Naïve Bayes* dan seleksi fitur *Chi-Square* yang diimplementasikan dalam bahasa pemrograman Python. Penelitian ini juga bertujuan untuk mengevaluasi pengaruh hasil klasifikasi dengan seleksi fitur dan tanpa seleksi fitur terhadap kinerja *Multinomial Naïve Bayes*. *Threshold Chi-Square* yang digunakan bervariasi dari 10% hingga 100%, dengan penambahan kelipatan 10% pada setiap iterasi, dan dilakukan validasi dengan menggunakan nilai *fold cross validation* = 10 pada setiap iterasi nilai *Threshold Chi-Square*.

3.1. Hasil Text Pre-Processing

Berdasarkan hasil penyeimbangan dan *labeling* data yang mencakup kelas positif dengan label 0 sebanyak 1536 data, dan kelas negatif sebanyak 1536 data, selanjutnya dilakukan *pre-processing text* dengan tujuan keseragaman, menghilangkan *noise data*, dan meningkatkan kinerja dari model yang dibangun. Hasil *text pre-processing* disajikan pada tabel 2.

Tabel 2. Hasil Setiap Tahapan *Text Pre-Processing*.

No.	Tahapan	Hasil Setiap Tahapan
1	Data awal	YANG MENANG BLM TENTU MULIA DAN YG KALAH BLM TENTU HINA!!!!!! DUNIA INI CM PERMAINAN HUKUMPUN BS DIMAINKAN DAN DIAKHIRATLAH
2	<i>Cleaning</i>	yang menang blm tentu mulia dan yg kalah blm tentu hina dunia ini cm permainan hukumpun bs dimainkan dan diakhiratlah nanti akan diadili seadiladilnya
3	Normalisasi	yang menang belum tentu mulia dan yang kalah belum tentu hina dunia ini cuma permainan hukumpun bisa dimainkan dan diakhiratlah nanti akan diadili seadiladilnya
5	<i>Tokenization</i>	['yang', 'menang', 'belum', 'tentu', 'mulia', 'dan', 'yang', 'kalah', 'belum', 'tentu', 'hina', 'dunia', 'ini', 'cuma', 'permainan', 'hukumpun', 'bisa', 'dimainkan', 'dan', 'diakhiratlah', 'akan', 'diadili', 'seadiladilnya']
4	<i>Filtering</i>	['menang', 'mulia', 'kalah', 'hina', 'dunia', 'permainan', 'hukumpun', 'dimainkan', 'diakhiratlah', 'diadili', 'seadiladilnya']
6	<i>Stemming</i>	['menang', 'mulia', 'kalah', 'hina', 'dunia', 'main', 'hukum', 'dimainkan', 'akhirat', 'adil', 'adil']

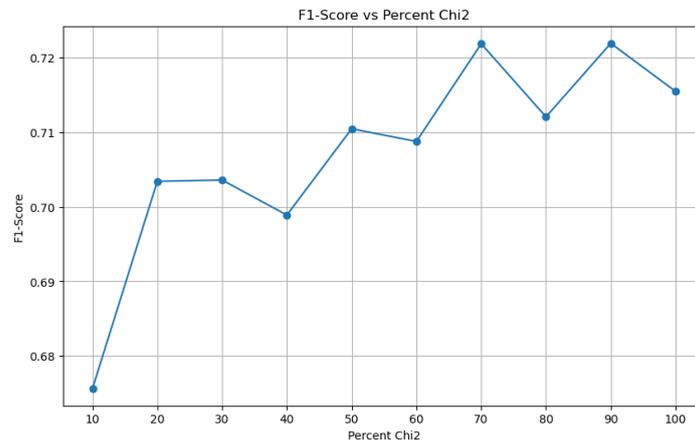
3.2. Hasil Implementasi Algoritma *Multinomial Naive Bayes* dan Seleksi Fitur *Chi-Square*

Algoritma *Multinomial Naïve Bayes* diimplementasikan dengan data latih sebesar 80% dan 20% sebagai data uji dengan ekstraksi fitur menggunakan TF-IDF yang uji coba menggunakan 10 nilai *Threshold Chi-Square* yang berbeda dan untuk validasi model menggunakan nilai *fold* = 10. Berdasarkan skenario tersebut, diperoleh kinerja model yang mencakup akurasi, presisi, recall, dan *F1-score* pada tabel 3.

Tabel 3. Pengujian Model dengan Variasi *Threshold Chi-Square*.

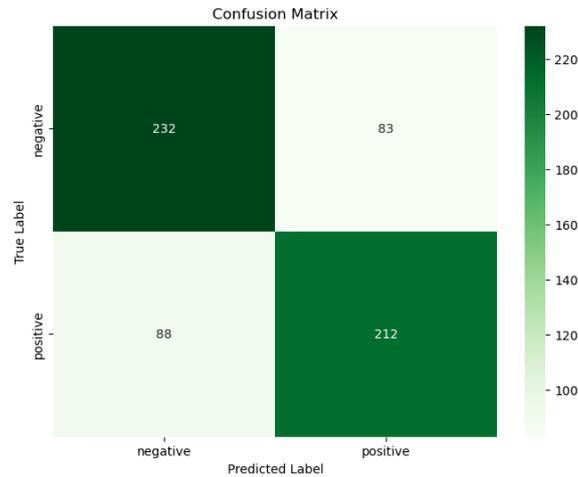
<i>Chi-square threshold</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
10%	0.6780487804878049	0.6809462225099708	0.6780487804878049	0.6756365362342023
20%	0.7040650406504065	0.7045980787925207	0.7040650406504065	0.7034258281742602
30%	0.7040650406504065	0.7043436794344943	0.7040650406504065	0.7035989346147757
40%	0.6991869918699187	0.6992330245602107	0.6991869918699187	0.6988968910728157
50%	0.7105691056910569	0.7105114213453788	0.7105691056910569	0.7104725735650723
60%	0.7089430894308943	0.7089142197263897	0.7089430894308943	0.7087766166143824
70%	0.7219512195121951	0.7218995452666391	0.7219512195121951	0.7218776275634126
80%	0.7121951219512195	0.7121451760592691	0.7121951219512195	0.712077787716335
90%	0.7219512195121951	0.7219314974996411	0.7219512195121951	0.7219394492863624
100%	0.7154471544715447	0.7157167929393241	0.7154471544715447	0.7154892925009808

Berdasarkan hasil pengujian pada tabel 9, nilai *threshold Chi-Square* yang lebih tinggi menyebabkan kriteria seleksi fitur menjadi lebih ketat dengan hanya mempertimbangkan fitur-fitur yang sangat signifikan dalam pembentukan model. Dengan *threshold* sebesar 70% dan 90%, model dapat mengakses fitur-fitur yang memiliki korelasi sangat signifikan dengan label sentimen, meningkatkan kinerja model dalam memprediksi sentimen dari teks ulasan.



Gambar 3. Grafik *F1-Score* pada Berbagai Nilai *Threshold Chi-Square*.

Berdasarkan grafik pada gambar 3, penggunaan seleksi fitur *Chi-Square* dapat memperbaiki performa sistem dengan menyaring kata-kata tidak berkorelasi. Performa sistem cenderung meningkat dengan lebih banyak fitur, tetapi terlalu sedikit fitur dapat menurunkan performa. Pada penggunaan fitur sebesar 70% dan 90%, terjadi peningkatan performa signifikan meskipun terdapat kata-kata tidak baku atau bahasa asing. Memilih semua fitur (100%) dapat menyebabkan *overfitting* dan redundansi sehingga menurunkan performansi model, sedangkan seleksi fitur 70% dan 90% mampu mempertahankan fitur paling informatif dan meningkatkan akurasi tanpa fitur yang tidak relevan. Ulasan yang tidak dapat diklasifikasikan terjadi karena pengurangan fitur, yang dapat mengakibatkan penghapusan ulasan dari data uji coba.



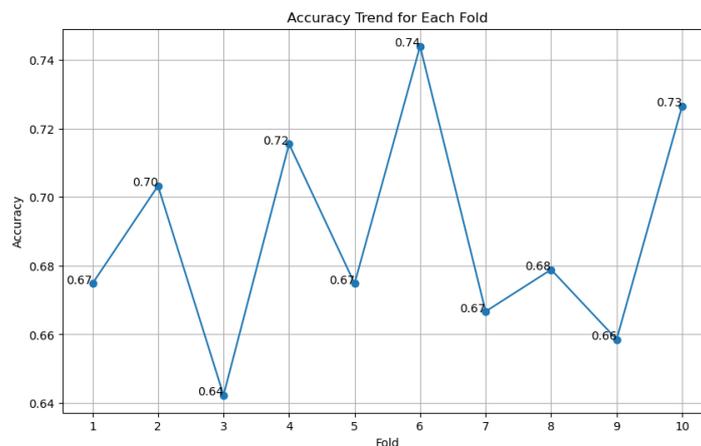
Gambar 4. Visualisasi *Confusion Matrix* Kinerja Model.

Pada gambar 4, kinerja model digambarkan dalam *confusion matrix* dimana model berhasil memprediksi dengan tepat sebesar 232 data pada label negatif dan sebanyak 212 data berlabel positif. Namun sebanyak 83 data yang semestinya berlabel negatif namun diprediksi positif dan sebanyak 88 data yang seharusnya berlabel positif namun diprediksi negatif oleh model. Gambaran tersebut juga tertuang dalam laporan klasifikasi kinerja model yang ditunjukkan pada tabel 4.

Tabel 4. Laporan Klasifikasi Kinerja Model.

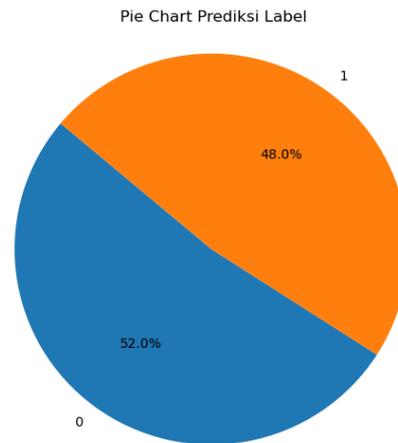
	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
negatif	0.72	0.74	0.73	315
positif	0.72	0.71	0.71	300
<i>accuracy</i>			0.72	615
<i>macro avg</i>	0.72	0.72	0.72	615
<i>weighted avg</i>	0.72	0.72	0.72	615

Berdasarkan tabel 4, model berhasil memprediksi data berlabel negatif dengan akurasi 73% dan berhasil memprediksi data berlabel positif sebesar 71% sehingga model yang dibangun memiliki kinerja yang cukup baik dalam menganalisa sentimen/emosi dari sebuah komentar.



Gambar 5. Visualisasi Grafik Nilai Akurasi Setiap Iterasi *Fold*.

Berdasarkan gambar 5, validasi model dengan nilai *fold* sebanyak 10 diperoleh informasi bahwa nilai *accuracy* tertinggi yakni pada *fold* ke-7 dengan akurasi 74% dan rata-rata akurasi dari seluruh iterasi *fold* sebesar 69% dan minimum akurasinya adalah 64% pada *fold* ke 3. Hal ini menunjukkan bahwa meskipun terdapat variasi dalam kinerja model di setiap iterasi *fold*-nya, namun secara keseluruhan model mampu mencapai tingkat akurasi akhir yang konsisten dan stabil di semua aspek pengukuran.



Gambar 6. Hasil Prediksi Perbandingan Sentimen Masyarakat.

Berdasarkan kasus yang diangkat dan hasil pengujian, model yang dibangun menunjukkan bahwa respon masyarakat terutama di platform YouTube memberikan respons negatif terhadap putusan sengketa Pilpres 2024. Hal itu terlihat pada gambar 6, bahwa label negatif (0) sebesar 52%, sedangkan label positif (1) hanya 48% yang menunjukkan mayoritas masyarakat merasa tidak puas terhadap keputusan MK. Hal ini dapat menjadi masukan untuk pemerintah agar meningkatkan kedaulatan hukum dan demokrasi di Indonesia.

4. Kesimpulan

Berdasarkan hasil pengujian dan rumusan masalah yang telah ditetapkan, dapat disimpulkan bahwa implementasi Algoritma *Multinomial Naïve Bayes* dengan seleksi fitur *Chi-Square* mampu memberikan kinerja yang cukup optimal dalam menganalisa dan memprediksi sentimen pada komentar masyarakat. Hal ini dibuktikan dengan kinerja model yang optimal dengan nilai *Threshold Chi-Square* = 70% dan 90% dapat memberikan akurasi, presisi, *recall*, dan *f1-score* yang stabil di angka 72%. Sedangkan kinerja model tanpa seleksi fitur atau nilai *Threshold* = 100% menghasilkan akurasi 71.5% sehingga dengan komposisi seleksi fitur yang sesuai dengan kondisi data dapat memengaruhi kinerja model. Hasil tersebut diuji dengan perlakuan yang sama yakni 80% data latih dan 20% data uji dengan validasi nilai *k-fold* = 10. *K-Fold cross validation* memberikan nilai akurasi tertinggi pada *fold* ke-7 sebesar 74% dan rata-rata sebesar 69% yang menunjukkan bahwa model secara konsisten mencapai kinerja yang cukup baik. Analisis sentimen menggunakan model yang dibangun ini memberikan informasi bahwa masyarakat Indonesia pada platform youtube merasa kurang puas terhadap Putusan MK terkait Sengketa Pilpres 2024 yang ditunjukkan dengan rasio sentimen negatif sebesar 52% dibandingkan rasio sentimen positif yang hanya sebesar 48%. Penelitian selanjutnya diharapkan untuk menggunakan metode *deep learning* dan seleksi fitur yang berbeda serta melakukan *tuning hyperparameter* untuk optimalisasi kinerja model. Selain itu, dilakukan pertimbangan untuk menggunakan emoji, bahasa daerah campuran, dan bahasa sarkas sebagai indikator dalam menganalisa sentimen.

Daftar Pustaka

- [1] S. DPR RI, "Putusan MK Jadi Momentum Seluruh Pihak Bersatu Membangun Bangsa," Dpr.go.id, 2020. Tersedia: <https://www.dpr.go.id/berita/detail/id/49311/t/Putusan+MK+Jadi+Momentum+Seluruh+Pihak+Bersatu+Membangun+Bangsa> [Diakses: 7 Mei 2024].
- [2] N.A.R. Bria, and A. Witanti, Analisis Sentimen Masyarakat Indonesia Menggunakan Algoritma Support Vector Machine Tentang Pilpres. *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol 7, no. 6, pp.3328-3333, 2024.
- [3] A. Habiba, R.R. Isnanto, and J.E. Suseno, Pemilihan Fitur Chi Square Pada Algoritma Naïve Bayes dan Pengaruhnya Terhadap Analisis Sentimen Masyarakat Indonesia Tentang Pembelajaran Tatap Muka Pada Masa Pandemi Covid-19. *vol, 12*, pp.111-116, 2023.
- [4] A.P. Permana, T. Chamidy, dan C. Crysdiyan, "Klasifikasi ulasan fasilitas publik menggunakan metode Naïve Bayes dengan seleksi fitur Chi-square", *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 8, no. 2, pp.112-124, 2023.
- [5] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, p. 406, Apr. 2021, doi: 10.30865/mib.v5i2.2835.
- [6] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, "The impact of features extraction on the sentiment analysis," in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 341–348. doi: 10.1016/j.procs.2019.05.008.
- [7] A. A. Farisi, Y. Sibaroni, and S. A. Faraby, "Sentiment analysis on hotel reviews using *Multinomial Naïve Bayes* classifier," *Journal of Physics: Conference Series*, vol. 1192, p. 012024, Mar. 2019, doi: 10.1088/1742-6596/1192/1/012024.
- [8] Nurainun, E. Haerani, F. Syafria, and L. Oktavia, "Penerapan Algoritma Naïve Bayes Classifier Dalam Klasifikasi Status Gizi Balita dengan Pengujian K-Fold Cross Validation," *Journal of Computer System and Informatics/Journal of Computer System and Informatics (JoSYC)*, vol. 4, no. 3, pp. 578–586, May 2023, doi: <https://doi.org/10.47065/josyc.v4i3.3414>.

Halaman ini sengaja dibiarkan kosong