

**ANALISIS SENTIMEN DATA TWITTER
CALON PRESIDEN 2024 MENGGUNAKAN ALGORITMA
NAÏVE BAYES CLASSIFIER DENGAN METODE
K-FOLD CROSS VALIDATION**

SKRIPSI

*Diajukan sebagai salah satu persyaratan untuk memperoleh gelar
Sarjana Statistika*



Oleh
ALDI PRAJELA
NIM 19337038

**PROGRAM STUDI SARJANA STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI PADANG
2024**

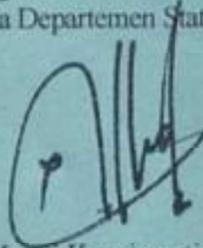
PERSETUJUAN SKRIPSI

ANALISIS SENTIMEN DATA TWITTER CALON PRESIDEN 2024 MENGGUNAKAN ALGORITMA *NAÏVE BAYES CLASSIFIER* DENGAN METODE *K-FOLD CROSS VALIDATION*

Nama : Aldi Prajela
NIM : 19337038
Program Studi : SI Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

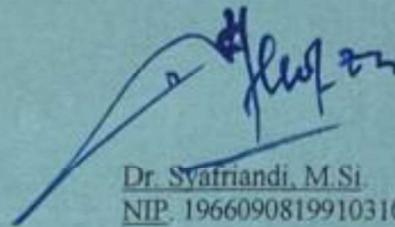
Padang, April 2024

Mengetahui:
Ketua Departemen Statistika



Dr. Yenni Kurniawati, S.Si., M.Si
NIP. 198402232010122005

Disetujui Oleh:
Pembimbing



Dr. Syafriandi, M.Si
NIP. 196609081991031003

PENGESAHAN LULUS UJIAN SKRIPSI

Nama : Aldi Prajela
NIM : 19337038
Program Studi : SI Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

ANALISIS SENTIMEN DATA TWITTER CALON PRESIDEN 2024 MENGGUNAKAN ALGORITMA *NAÏVE BAYES CLASSIFIER* DENGAN METODE *K-FOLD CROSS VALIDATION*

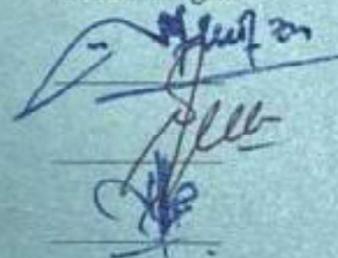
Dinyatakan lulus setelah dipertahankan di depan Tim Penguji Skripsi
Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Padang

Padang, April 2024

Tim Penguji

	Nama
Ketua	: Dr. Syafriandi, M.Si
Anggota	: Dr. Dony Permana, M.Si
Anggota	: Dina Fitria, S.Pd., M.Si

Tanda Tangan



SURAT PERNYATAAN TIDAK PLAGIAT

Saya yang bertanda tangan di bawah ini:

Nama : Aldi Prajela

NIM : 19337038

Program Studi : S1 Statistika

Departemen : Statistika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan bahwa, skripsi saya dengan judul “**Analisis sentimen data Twitter calon Presiden 2024 menggunakan *Algoritma Naive Bayes Classifier* dengan metode *K-fold cross validaton*”** adalah benar merupakan hasil karya saya dan bukan merupakan plagiat dari karya orang lain atau pengutipan dengan cara-cara yang tidak sesuai dengan etika yang berlaku dalam tradisi keilmuan.

Apabila suatu saat terbukti saya melakukan plagiat maka saya bersedia diproses dan menerima sanksi akademis maupun hukum sesuai dengan hukum dan ketentuan yang berlaku, baik di institusi UNP maupun di masyarakat dan negara.

Demikian pernyataan ini saya buat dengan penuh kesadaran rasa tanggung jawab sebagai anggota masyarakat ilmiah.

Diketahui oleh,
Kepala Departemen Statistika,



Dr. Yenni Kurniawati, S.Si., M.Si
NIP. 198402232010122005

Saya yang menyatakan,



Aldi Prajela
NIM. 19337038

Analisis Sentimen Data Twitter Calon Presiden 2024 Menggunakan Algoritma *Naïve Bayes Classifier* dengan Metode *K-Fold Cross Validation*

Aldi Prajela

ABSTRAK

Pemilihan Presiden 2024 adalah sebuah proses memilih Presiden Republik Indonesia masa bakti 2024–2029. Pemilihan ini merupakan pilpres yang ke-8 di Indonesia. Menjelang dimulainya Pilpres, calon presiden menjadi perbincangan yang sering dibicarakan sehari-hari. Opini-opini positif dan negatif yang disampaikan oleh masyarakat pada media sosial twitter mengenai Pilpres dan masing-masing calonnya sering menjadi *trending*, hal ini menandakan bahwa banyak masyarakat Indonesia yang menyuarakan opininya di *twitter*. Analisis yang digunakan untuk menyerap *tweet* tersebut menjadi sebuah informasi adalah analisis sentimen.

Metode yang digunakan dalam pengklasifikasian sentimen yaitu *Naïve Bayes Classifier* dengan metode *K-Fold Cross Validation* untuk mendapatkan hasil akurasi yang terbaik. Sebelum melakukan tahap pemodelan, *tweet* dikumpulkan melalui teknik *crawling* dan melalui tahapan seperti *text preprocessing*, pelabelan berbasis *lexicon based*, serta pembobotan data. Model yang dihasilkan dievaluasi menggunakan *confusion matrix*.

Hasil penelitian menunjukkan bahwa hasil klasifikasi NBC terhadap sentimen masyarakat kepada masing-masing Capres yaitu, Anies mendapatkan 83% *tweet* positif dan 17% *tweet* negatif dari 2148 *tweet*, Prabowo Subianto mendapatkan 77% *tweet* positif dan 23% *tweet* negatif dari 1870 *tweet*, sedangkan Ganjar Pranowo mendapatkan 74% *tweet* positif dan 26% *tweet* negatif dari 1557 *tweet*. Setelah dilakukan *confusion matrix* untuk setiap *keyword* maka didapatkan hasil untuk *keyword* Anies memiliki nilai tertinggi pada *iteration* ke-6 yaitu nilai *accuracy* sebesar 80%, dan nilai *recall* sebesar 95%. Selanjutnya Prabowo memiliki nilai tertinggi pada *iteration* ke-10 yaitu nilai *accuracy* sebesar 80%, dan nilai *recall* sebesar 90%. Sedangkan Ganjar memiliki nilai tertinggi pada *iteration* 8 yaitu *accuracy* sebesar 85%, dan nilai *recall* sebesar 94%.

Kata Kunci: Analisis Sentimen, Calon Presiden, *K-Fold*, *Naïve Bayes*

Twitter Data Sentimen Analysis 2024 Presidential Candidate Using Algorithm Naïve Bayes Classifier By Methods K-Fold Cross Validation

Aldi Prajela

ABSTRACT

The 2024 Presidential Election is a process to elect the President of the Republic of Indonesia for the 2024–2029 term. This election is the 8th presidential election in Indonesia. Towards the start of the presidential election, presidential candidates have become a topic of conversation that is often discussed on a daily basis. The positive and negative opinions expressed by the public on Twitter social media regarding the Presidential Election and each candidate often become trending, this indicates that many Indonesians voice their opinions on Twitter. The analysis used to absorb these tweets into information is sentiment analysis. The method used in sentiment classification is the Naïve Bayes Classifier with the K-Fold Cross Validation method to get the best accuracy results. Before carrying out the modeling stage, tweets were collected using crawling techniques and through stages such as text preprocessing, lexicon based labeling, and data weighting. The resulting model was evaluated using a confusion matrix.

The results of the research show that the results of NBC's classification of public sentiment towards each presidential candidate, they are, Anies got 83% positive tweets and 17% negative tweets from 2148 tweets, Prabowo Subianto got 77% positive tweets and 23% negative tweets from 1870 tweets, while Ganjar Pranowo received 74% positive tweets and 26% negative tweets from 1557 tweets. After carrying out the confusion matrix for each keyword, the results obtained for the keyword Anies had the highest value in the 6th iteration, with accuracy value 80%, and recall value of 95%. Furthermore, Prabowo had the highest score in the 10th iteration, with accuracy value 80%, and recall value 90%. Meanwhile, Ganjar had the highest score in iteration 8, with accuracy 85%, and recall value 94%.

Keywords: K-Fold, Naïve Bayes, Presidential Candidates, Sentiment Analysis.

KATA PENGANTAR

Alhamdulillah, rasa syukur penulis ucapkan kehadiran Allah SWT atas berkat rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul “**Analisis Sentimen Data Twitter Calon Presiden 2024 Menggunakan Algoritma Naïve Bayes Classifier Dengan Metode K-Fold Cross Validation**”. Penulisan Skripsi ini disusun untuk memenuhi salah satu syarat dalam menyelesaikan pendidikan pada Program Studi Statistika (S1) Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang. Dalam menyelesaikan Skripsi ini, tidak terlepas dari bimbingan dan bantuan dari berbagai pihak. Penulis menyadari bahwa skripsi ini memiliki keterbatasan sesuai dengan ilmu yang dimiliki, oleh sebab itu penulis menerima saran dan kritikan dari pembaca demi kesempurnaan isi skripsi ini penulis mengucapkan terima kasih kepada.

1. Ibu Dr. Yenni Kurniawati., M.Si., Kepala Departemen Statistika dan Koordinator Program Studi S1 Statistika FMIPA UNP.
2. Bapak Dr. Dony Permana, M.Si., dan Ibu Dina Fitria, M.Si., Dosen Penguji Skripsi yang telah memberikan masukan serta saran pada skripsi ini.
3. Bapak Dr. Syafriandi., M.Si., Dosen Pembimbing Skripsi dan Penasehat Akademik yang telah membimbing dan memberikan arahan selama masa perkuliahan dan penyusunan skripsi ini.
4. Bapak dan Ibu Staf Pengajar dan Tenaga Kependidikan Departemen Statistika Fakultas Matematika Ilmu Pengetahuan Alam Universitas Negeri Padang yang telah membantu penulis selama menimba ilmu di Departemen

Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang.

5. Teristimewa kepada kedua orang tua dan kedua saudara penulis yang telah memberikan kasih sayang, semangat, nasehat, dukungan dan doa selama masa perkuliahan dan dalam menyelesaikan skripsi ini.
6. Semua teman-teman yang selalu memberikan semangat serta dukungan kepada penulis selama menempuh perkuliahan di Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang.

Semoga semua kebaikan dan ketulusan dibalas oleh Allah SWT sebagai amal ibadah. Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna dan masih terdapat kekurangan didalamnya. Oleh karena itu, penulis mengharapkan masukan dan saran yang membangun dari pembaca. Semoga penulisan skripsi ini dapat bermanfaat bagi seluruh pihak yang memerlukannya.

Padang, Februari 2024

Penulis

Aldi Prajela

DAFTAR ISI

ABSTRAK	i
KATA PENGANTAR	iii
DAFTAR ISI	v
DAFTAR TABEL	vii
DAFTAR GAMBAR	vii
DAFTAR LAMPIRAN	ix
BAB I PENDAHULUAN	1
A. Latar Belakang	1
B. Batasan Masalah.....	6
C. Rumusan Masalah	6
D. Tujuan Penelitian	6
E. Manfaat Penelitian	7
BAB II KAJIAN TEORI	8
A. Pemilihan Umum	8
1. Pemilihan Presiden.....	9
2. Bakal Calon Presiden	10
a. Anies Rasyide Baswedan	10
b. Ganjar Pranowo.....	11
c. Prabowo Subianto	13
B. Twitter	13
C. Analisis Sentimen.....	16
D. <i>Pre-proccesing Data</i>	17
1. <i>Cleansing</i>	17
2. <i>Normalizing</i>	17
3. <i>Case Folding</i>	18
4. <i>Tokenizing</i>	18
5. <i>Stemming</i>	18
6. <i>Stopword Removal</i>	19
E. Pembobotan <i>Term Frequency – Inverse Document Frequency</i>	19
F. Metode <i>Lexicon Based</i>	20

G. Metode <i>Cross Validation</i>	24
H. Klasifikasi	24
I. <i>Naïve Bayes Classifier</i>	24
J. Ketetapan Klasifikasi	27
K. <i>Word Cloud</i>	30
BAB III METODOLOGI PENELITIAN	31
A. Jenis Penelitian.....	31
B. Jenis Dan Sumber Data Penelitian	31
C. Variabel Penelitian	31
D. Metode Pengumpulan Data	32
E. Teknik Analisis Data.....	34
F. Evaluasi Model.....	35
G. Diagram Alur Penelitian	36
BAB IV HASIL DAN PEMBAHASAN	37
A. Hasil Penelitian	37
1. <i>Preprocessing</i>	37
2. TF-IDF	41
3. Pelabelan <i>Lexicon Based</i>	42
4. <i>Word Cloud</i>	45
5. <i>K-fold Cross Validation</i>	48
6. Klasifikasi <i>Naïve Bayes Classifier</i>	49
7. Evaluasi Model.....	60
8. Pembahasan.....	63
BAB V PENUTUPAN	
A. Kesimpulan	65
B. Saran.....	66

DAFTAR TABEL

Tabel	Halaman
1. Confusion Matrix	29
2. Kriteria Kualitas Akurasi	29
3. Kriteria Kualitas <i>Recall</i>	30
4. Variabel Untuk Klasifikasi Sentimen.....	32
5. Contoh Ulasan Twitter	37
6. Contoh <i>Cleansing Data</i>	38
7. Contoh <i>Case Folding</i>	38
8. Contoh <i>Tokenizing</i>	39
9. Contoh <i>Normalizing</i>	40
10. Contoh <i>Stopword</i>	40
11. Contoh <i>Stemming</i>	41
12. Contoh Perhitungan TF-IDF	42
13. Contoh Hasil Data Menggunakan VADER.....	44
14. Hasil Pelabelan <i>Tweet</i>	44
15. Hasil Pembagian Data <i>Training</i> dan <i>Testing</i>	49
16. Frekuensi <i>Term</i> Pada Setiap Kelas Sentimen.....	49
17. Hasil perhitungan Peluang	49
18. Frekuensi <i>Term</i> Pada Setiap Kelas Sentimen.....	52
19. Hasil perhitungan Peluang	54
20. Frekuensi <i>Term</i> Pada Setiap Kelas Sentimen.....	56
21. Hasil perhitungan Peluang	57
22. <i>Confusion Matrix</i> Hasil Prediksi Model NBC	59
23. Hasil Evaluasi Model	62

DAFTAR GAMBAR

Gambar	Halaman
1. Ilustrasi Metode <i>K-fold Cross Validation</i>	23
2. Contoh Visualisasi Kata <i>Word Cloud</i>	30
3. Diagram alur penelitian.....	36
4. <i>Word Cloud</i> sentimen Positif dan Negatif Anies	46
5. <i>Word Cloud</i> sentimen Positif dan Negatif Prabowo	47
6. <i>Word Cloud</i> sentimen Positif dan Negatif Ganjar.....	48
7. Diagram Hasil Klasifikasi NBC Anies	52
8. Diagram Hasil Klasifikasi NBC Prabowo	55
9. Diagram Hasil Klasifikasi NBC Ganjar	59
10. Diagram Hasil Klasifikasi NBC	59

DAFTAR LAMPIRAN

Lampiran	Halaman
1. <i>Crawling Data</i>	73
2. Proses <i>Pre-Processing</i>	73
3. Proses Pelabelan Data Menggunakan Metode <i>Lexicon Based</i>	76
4. Proses Pembobotan TF-IDF.....	77
5. Proses Klasifikasi NBC menggunakan Metode <i>K-Fold Cross Validation</i> dan Evaluasi <i>Confusion Matrix</i>	79
6. Hasil Klasifikasi Model NBC terhadap data <i>training</i> Anies.....	81
7. Hasil Klasifikasi Model NBC terhadap data <i>training</i> Prabowo.....	108
8. Hasil Klasifikasi Model NBC terhadap data <i>training</i> Ganjar	125

BAB I PENDAHULUAN

A. Latar Belakang Masalah

Indonesia adalah salah satu negara yang menganut sistem demokrasi. Sebagai negara yang menganut sistem demokrasi, penting bagi warga negara Indonesia memiliki sebuah proses untuk memilih orang yang dapat mengisi jabatan-jabatan politik tertentu. Proses tersebut dikenal sebagai Pemilihan Umum (Pemilu). Di Indonesia diketahui bahwa pemilu diselenggarakan secara periodik yaitu selama lima tahun sekali. Pada tahun 2024 diselenggarakan kembali pemilu, salah satunya adalah Pemilihan Presiden (Pilpres). Gusti (2019), menyatakan bahwa masyarakat juga menjadi sasaran elit politik, dimana suara mereka merupakan penentu keberlangsungan arah politik untuk lima tahun kedepan.

Pilpres 2024 adalah sebuah proses demokrasi untuk memilih Presiden Republik Indonesia masa bakti 2024–2029 yang akan dilaksanakan pada hari Rabu, tanggal 14 Februari 2024. Pemilihan ini merupakan Pilpres yang ke-8 di Indonesia. Undang-Undang RI Nomor 7 Tahun 2017 tentang Pemilihan Umum yang menjadi pedoman penyelenggaraan Pilpres dan Wakil Presiden tahun 2019, telah mengatur mekanisme prosedur dan persyaratan menjadi calon Presiden dan Wakil Presiden. Adapun untuk calon Presiden dan Wakil Presiden harus memenuhi batas ambang pencalonan. Pada pasal 222 Undang-Undang Republik Indonesia Nomor 7 Tahun 2017 disebutkan bahwa “Pasangan Calon diusulkan oleh Partai Politik atau Gabungan Partai Politik Peserta Pemilu yang memenuhi persyaratan perolehan kursi paling sedikit 20% (dua puluh persen) dari jumlah

kursi DPR atau memperoleh 25% (dua puluh lima persen) dari suara sah secara nasional pada Pemilu anggota DPR sebelumnya”.

Memasuki era sebelum Pilpres, masyarakat turut aktif menyampaikan opininya mengenai berbagai hal yang berkaitan dengan hal tersebut. Salah satu topik yang tidak luput dari pembahasan masyarakat adalah opini tentang setiap calon yang mendaftar pada Pilpres 2024. Adapun tokoh publik yang mencalonkan diri sebagai calon Presiden Republik Indonesia untuk periode 2024-2029 adalah Anies Baswedan, Ganjar Pranowo, dan Prabowo Subianto. Menjelang dimulainya Pilpres, calon presiden menjadi perbincangan yang sering dibicarakan dalam kehidupan sehari-hari. Namun saat ini, masyarakat cenderung menyampaikan opininya pada *platform* media sosial seperti Instagram, Threads, Facebook, dan Twitter. Masyarakat juga menjadi sasaran elit politik, dimana suara mereka merupakan penentu keberlangsungan arah politik untuk lima tahun kedepan. Opini-opini positif serta negatif yang disampaikan oleh masyarakat pada media sosial sangat beragam dan tergantung pada berbagai faktor, termasuk latar belakang politik, keyakinan, nilai-nilai, dan pengalaman pribadi. Pada media sosial twitter, topik mengenai Pilpres dan masing-masing calonnya sering menjadi *trending*, hal ini menandakan bahwa banyak masyarakat Indonesia yang menyuarkan opininya di *twitter*.

Twitter merupakan media sosial yang dibuat oleh Jack Dorsey pada tahun 2006. Pada tahun 2019 berdasarkan *press-release Twitter* terdapat 500 juta *tweet* atau kicauan oleh pengguna *twitter* per harinya. Sebanyak 500 juta *tweet* digunakan untuk mem-post hal tentang diri pengguna dan berbagi informasi. Isi

tweet juga dapat mengekspresikan perasaan (Mahbubah, 2019). Data opini seperti *tweet* yang disampaikan oleh masyarakat tersebut dapat diserap menjadi sebuah informasi apabila dilakukan analisis untuk memilah apakah opini tersebut bersifat positif ataupun negatif. Analisis yang digunakan untuk menyerap *tweet* tersebut menjadi sebuah informasi adalah analisis sentimen.

Menurut Sabily (2019), analisis sentimen merupakan suatu metode yang menganalisis suatu pendapat, opini, evaluasi, sentimen, sikap, atau penilaian seseorang terhadap seorang individu, produk, masalah, organisasi, peristiwa, atau topik. Berdasarkan penelitian Mahbubah (2019), terdapat beberapa metode yang dapat digunakan pada analisis sentimen, diantaranya adalah *Naïve Bayes Classifier*, *Decision Trees*, dan *Support Vector Machines*. Metode yang paling populer digunakan dalam pengklasifikasian dokumen sekarang ini adalah metode *Naïve Bayes Classifier*. Metode *Naïve Bayes Classifier* memiliki beberapa kelebihan antara lain sederhana, mudah dipahami, dan cocok untuk klasifikasi teks. Metode *Naïve Bayes classifier* mempunyai kecepatan dan akurasi yang tinggi ketika diaplikasikan dalam basis data yang besar dan data yang beragam. Meskipun demikian, pada data yang lebih kompleks masih terdapat kemungkinan terjadinya *overfitting*. Menurut Wildan (2023), banyak masalah yang terjadi pada data mining dikarenakan hasil yang diperoleh dipengaruhi oleh adanya ketidak seimbangan data sehingga terjadinya *overfitting* biasanya disebabkan oleh data yang berlebihan atau data yang kurang sehingga model yang dihasilkan tidak bisa menangkap pola dari data yang digunakan. *Overfitting* bisa merusak kinerja model dan mengurangi hasil akurasi dari sebuah model. Untuk

mengatasi kemungkinan *overfitting* pada klasifikasi *Naïve Bayes* dilakukan dengan *cross validation*.

Menurut Raschka (2018), *cross validation* berfungsi membagi data menjadi dua bagian yaitu data *training* dan data *testing* secara berulang, sehingga setiap data memiliki kesempatan yang sama untuk dibagi kembali menjadi data *training* dan *testing*. Raschka (2018) juga menyatakan bahwa salah satu teknik *cross validation* adalah *K-fold Cross Validation*. *K-fold Cross Validation* mengelompokkan data secara acak ke dalam k kelompok. Selanjutnya kelompok dibagi menjadi data *training* dan data *testing*. Hal ini dilakukan berulang kali sebanyak k dengan meninggalkan satu kelompok sebagai data *testing* di setiap perulangan *fold*.

Penelitian yang berkaitan dengan analisis sentimen telah dilakukan oleh Alif Firman (2023), pada penelitian tersebut menggunakan algoritma *Naive Bayes Classifier* (NBC) dalam mengklasifikasikan sentimen masyarakat mengenai Pilpres 2024 menggunakan *tools Rapid Miner*. Dimana ada empat kandidat yang telah diteliti yaitu, Anies mendapat 74% sentimen positif dan 26% sentimen negatif, disusul oleh Sandi yaitu 57% sentimen positif 43% sentimen negatif, Ganjar mendapat 53% sentimen positif 47% sentimen negatif dan Prabowo mendapat 32% sentimen positif 68% sentimen negatif. Penelitian yang serupa dilakukan oleh Asno (2023), mengenai pengklasifikasi sentimen pada proyeksi Pilpres 2024 menggunakan metode *Support Vector Machine*. Pada penelitian tersebut diperoleh hasil sentimen berdasarkan tiga dataset kandidat yang dipilih, yaitu Anies Baswedan 65,62%, Ganjar Pranowo 73,58%, dan Prabowo Subianto

66,34%. Hasil akurasi metode yang dimiliki oleh ketiga dataset yaitu Anies Baswedan 73%, Ganjar Pranowo 79% dan Prabowo Subianto 79%.

Penelitian lainnya dilakukan oleh Herfia & Jaka (2019), membahas pengklasifikasi sentimen Berita Kriminal Menggunakan *Naïve Bayes Classifier* (NBC) dengan metode *K-fold Cross Validation*. Hasil pengujian menyatakan bahwa NBC memiliki kemampuan dalam klasifikasi berita kriminal dengan nilai *precision* sebesar 98,53 %, nilai *recall* sebesar 98,44 % dan nilai *accuracy* sebesar 99,38 %.

Berdasarkan penjabaran di atas, dapat diketahui bahwa metode NBC dengan metode *K-fold Cross Validation* dapat memberikan akurasi yang tinggi dalam memodelkan klasifikasi sentimen. Oleh sebab itu, peneliti menggunakan algoritma NBC dengan metode *K-fold Cross Validation* dalam mengklasifikasikan sentimen pengguna *Twitter* mengenai calon Presiden 2024 dengan penelitian yang berjudul **“Analisis Sentimen Data *Twitter* Calon Presiden 2024 Menggunakan Algoritma *Naïve Bayes Classifier* dengan metode *K-fold Cross Validation*”**.

B. Batasan Masalah

Batasan masalah pada penelitian ini difokuskan pada pengguna aplikasi *Twitter*, yang memberikan opini dalam bahasa Indonesia terhadap Pilpres 2024 dimulai pada tanggal 1 Oktober 2023 - 30 November 2023.

C. Rumusan Masalah

Rumusan masalah dari penelitian ini adalah:

1. Bagaimana hasil klasifikasi NBC sentimen masyarakat pada sosial media twitter terhadap Pilpres 2024?
2. Bagaimana performa model klasifikasi Algoritma NBC dengan metode *K-fold Cross Validation* dalam analisis sentiment untuk mengklasifikasikan sentimen masyarakat mengenai Pilpres 2024.

D. Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Untuk mengetahui bagaimana hasil klasifikasi NBC sentimen masyarakat pada sosial media twitter mengenai Pilpres 2024.
2. Untuk mengetahui performa model klasifikasi Algoritma NBC dengan metode *K-fold Cross Validation* dalam analisis sentiment untuk mengklasifikasikan sentimen masyarakat mengenai Pilpres 2024.

E. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat kepada:

1. Bagi peneliti, sarana menambah ilmu pengetahuan serta wawasan mengenai sentimen persepektif masyarakat terhadap Pilpres 2024 pada

aplikasi *Twitter* menggunakan NBC dengan metode *K-fold Cross Validation*.

2. Bagi Masyarakat, hasil yang diperoleh dalam penelitian ini memungkinkan masyarakat untuk mengikuti seorang politisi pada media sosialnya khususnya pada aplikasi *Twitter*.
3. Bagi Pembaca, sebagai referensi bagi peneliti yang akan melakukan analisis sentimen dan NBC dengan metode *K-fold Cross Validation*