

# 1. Pendahuluan

## 1.1. Latar Belakang

Pemilihan Umum (Pemilu) Indonesia tahun 2024 telah menjadi salah satu topik menarik bagi para pengguna media sosial, khususnya *X* (atau *Twitter*). Pemilu ini merupakan pesta demokrasi terbesar di Indonesia yang meliputi pemilihan presiden dan wakil presiden oleh seluruh rakyat di Indonesia [1][2][3].

*X* atau *Twitter* merupakan salah satu *social media* yang dimanfaatkan untuk membagikan berbagai pendapat ke publik dengan fitur-fitur seperti *repost*, pengunggahan foto dan video, serta membagikannya ke jejaring sosial lainnya. “Orang-orang berbagi berita, mendiskusikan acara politik, dan memberikan komentar tentang peristiwa global tertentu. Oleh karena itu, media sosial juga kerap kali dipergunakan dalam kampanye politik untuk mempromosikan kegiatan sosial dan pembangunan, serta mengungkapkan sentimen tentang pemilu.” [4] khususnya dalam mengutarakan opini publik terhadap calon kandidat presiden Indonesia pada pemilihan umum 2024. Dengan demikian, penelitian ini dilakukan untuk menganalisis sentimen publik terhadap capres dan cawapres pada pemilihan umum 2024 dengan menggunakan data pada aplikasi *Twitter*.

Riset tentang analisis sentimen perihal Pemilu telah dilakukan oleh Rengga Asmara dkk [1] dengan Metode *Naïve Bayes* terhadap Pemilu 2019 yang mendapatkan hasil akurasi sebesar 62%, *recall* 45% dan presisi sebesar 45% dengan perbandingan 22,66% *positive statement* dan 39,39% *negative statement* sedangkan Lingga Aji Andika dkk [2] dengan metode yang sama menghasilkan akurasi sebesar 82,9% dengan perbandingan 34,5% *positive* dan 65,5% *negative statement*. Lalu jurnal karya Safitri Juanita [5] mendapatkan *recall* 72,3% dan presisi sebesar 81,6% dengan pembagian sentimen positif sebesar 18% dan negatifnya nya 52%. Karya Aprillia Rizki Adiati dkk [6] mendapatkan akurasi sebesar 78,03% dengan pembagian sentimen positif 53,8% dan negatif 46,13%. Lalu karya Mohammed Hafizh Al-Areef dkk [3] mendapatkan rata-rata akurasi 83,7%, presisi 86,3%, *recall* 94,3% dan *f1-score* 90,3% pada Pemilu 2024.

Nur Hayatin dkk [7] memperoleh akurasi sebesar 90,74% sedangkan Akhmad Muzaki dkk [8] menghasilkan akurasi sebesar 92,2% pada Pemilu 2019. Kemudian

Fadli Nurriziky dkk [9] menggunakan SVM mendapatkan akurasi sebesar 97% sedangkan NB mendapatkan 95%. Bobby Andika Putra dkk [10] dengan rata-rata akurasi SVM sebesar 89,24% sedangkan NB dengan rata-rata 83,79% dan Lisyana Damayanti dkk [11] memperoleh akurasi sebesar 90,75% pada pemilu 2024 dengan prediksi sentimen menggunakan data belum berlabel dilakukan.

Lalu untuk peneliti lainnya seperti Danar Wido Seno dkk [12] mendapatkan akurasi sebesar 92,5%, presisi antara 85,7% - 97,2% dan *recall* 78,2%-93,5% dengan menggunakan metode kombinasi *Lexicon Based* dan SVM pada Pemilu 2019. Okky Zoellanda dkk [13] mendapatkan akurasi sebesar 91,5% dengan memanfaatkan SVM pada pemilu yang sama, prediksi sentimen memanfaatkan data yang tidak berlabel belum dilakukan.

Meskipun metode yang digunakan untuk memprediksi sentimen berbeda-beda dengan tingkat akurasi yang berbeda-beda, namun hasil pelabelan prediksi sentimen yang dihasilkan dari beberapa penelitian cukup mirip. Metode yang digunakan berhasil memprediksi data dengan cukup baik. Dari penelitian yang sudah dijalankan sebelumnya, SVM merupakan metode yang tingkat akurasinya lebih baik dibandingkan NB, namun dari kedua penelitian tersebut belum ada yang melakukan uji prediksi terhadap data baru yang masih mentah. Di sisi lain, NB menghasilkan akurasi yang sedikit lebih rendah dibandingkan SVM dan *Lexicon-SVM*, namun dari beberapa penelitian yang telah dilakukan, metode ini berhasil memprediksi data mentah dan tetap menjaga performa model yang digunakan. Oleh karena itu, penggunaan metode kombinasi antara NB dan SVM dilakukan, guna meningkatkan akurasi yang dicapai dan juga memperoleh hasil penelitian terhadap data mentah baru.

Dari berbagai penelitian, jarang sekali ditemukan penelitian yang melakukan perbandingan antara sebelum dan sesudah keluarnya hasil pemilu. Dimana [1], [3], [5], [7]–[14] merupakan penelitian yang dibuat untuk menganalisa data-data *tweet* sebelum pemilu, sedangkan [2] dan [6] menganalisa data sesudah pemilu.

Metode yang dimanfaatkan dalam riset ini adalah *Naïve Bayes-Support Vector Machine*. NB adalah salah satu metode klasifikasi yang paling sederhana dan cepat. Metode ini memanfaatkan teorema *Bayes* untuk melakukan perhitungan probabilitas suatu data masukan tergolong ke dalam kelas tertentu [14]. *Support*

*Vector Machine* dikenal dengan model yang akurat dalam mengklasifikasikan data [15].

Dinamika adalah pola atau proses pertumbuhan, perubahan, atau perkembangan dalam suatu bidang tertentu, atau sistem keterkaitan yang saling mempengaruhi antara satu unsur dengan unsur lainnya. Jika salah satu unsur dalam sistem mengalami perubahan, maka unsur-unsur lainnya juga akan ikut berubah [16]. Dalam penelitian ini, telah dilakukan analisis dinamika sentimen publik terhadap calon presiden pada Pemilu 2024 dengan menggunakan data *Twitter*.

Berbeda dengan karya-karya sebelumnya, *dataset* ini telah diambil pada delapan belas waktu yang berbeda sehingga dapat menganalisa dinamika opini publik terhadap setiap capres dan cawapres untuk pemilihan umum 2024. Delapan belas waktu yang berbeda tersebut antara lain yakni dari masa pencalonan presiden dan wakil presiden, sela waktu antara pencalonan dengan masa kampanye, masa kampanye dimulai, sela waktu antara masa kampanye dengan debat pertama, debat pertama, sela waktu antara debat pertama dengan debat kedua, debat kedua, sela waktu antara debat kedua dengan debat ketiga, debat ketiga, sela waktu antara debat ketiga dengan debat keempat, debat keempat, sela waktu antara debat keempat dengan debat kelima, debat kelima, sela waktu antara debat kelima dengan hari pemilihan umum, hari pemilihan umum, perhitungan suara hasil pemilu dimulai, sela antara perhitungan suara, dan hasil pemilihan umum [17].

Data *Twitter* telah diambil dengan menggunakan *tools crawling Tweet-Harvest* karya Helmi Satria. *Dataset* diambil dengan kata kunci berikut: “Anies-Cak Imin”, “Prabowo-Gibran”, “Ganjar-Mahfud MD” serta "El Chef", "El Slepet", "El Gemoy", "El Sulfat", "El Chudai", dan "El Professor" pada waktu debat.

## 1.2. Perumusan Masalah

Adapun rumusan masalah dalam pelaksanaan riset ini, diantaranya:

1. Bagaimana dinamika sentimen publik terhadap masing-masing capres dan cawapres pada Pemilu Indonesia tahun 2024 pada beberapa periode waktu dan pola seperti apa yang akan terbentuk ketika acara resmi KPU diselenggarakan dengan menggunakan *Naïve Bayes-Support Vector Machine*?

2. Bagaimana akurasi, presisi, *recall* dan *f1-score* dari model yang dirancang dalam melakukan sentimen opini publik pada Pemilu Indonesia tahun 2024?
3. Apakah pendapat publik di media sosial (khususnya *Twitter*) selaras dengan hasil Pemilu Indonesia tahun 2024?

### 1.3. Tujuan

Adapun tujuan dalam pelaksanaan riset ini, diantaranya:

1. Untuk mengetahui dinamika sentimen publik terhadap masing-masing capres dan cawapres pada Pemilu Indonesia tahun 2024 pada beberapa periode waktu dan pola yang terbentuk ketika acara resmi KPU diselenggarakan dengan menggunakan *Naïve Bayes-Support Vector Machine*.
2. Untuk mengetahui akurasi, presisi, *recall* dan *f1-score* dari model yang dirancang dalam melakukan sentimen opini publik pada Pemilu Indonesia tahun 2024.
3. Untuk mengetahui apakah pendapat publik di media sosial (khususnya *Twitter*) selaras atau tidak dengan hasil Pemilu Indonesia tahun 2024.

### 1.4. Batasan Masalah

Adapun batasan yang perlu diperhatikan dalam penelitian ini terinci sebagai berikut:

1. *Dataset* diambil menggunakan kata kunci tertentu yakni: “Anies-Cak Imin”, “Prabowo-Gibran”, “Ganjar-Mahfud MD” serta "El Chef", "El Slepét", "El Gemoy", "El Sulfat", "El Chudai", dan "El Professor".
2. *Dataset* diambil dengan batasan waktu yang diawali pada tanggal 19 Oktober 2023 atau tanggal pendaftaran Paslon 01 dan Paslon 03 ke KPU dan diakhiri pada tanggal 20 Maret 2024 yakni tanggal keluarnya hasil pemilu oleh KPU.
3. Penelitian ini menggunakan metode *NB-SVM* sebagai model utama dan *Multinomial NB*, *Gaussian NB*, *SVM* dan *Kernel Linear* sebagai metode pembandingan untuk nilai akurasi, presisi, *recall* dan *f1-score*.

## 1.5. Sistematika Penulisan

Sistematika Penulisan yang digunakan dalam tugas akhir ini meliputi beberapa hal sebagai berikut:

1. Studi Literatur

Yang bertujuan untuk memahami konsep metode pendukung penyelesaian masalah dengan mempelajari konsep dan teori *Naïve Bayes* dan *Support Vector Machine*, serta meriset referensi yang berasal melalui *textbook*, *paper/jurnal* atau dokumen lain yang dapat mendukung penyelesaian permasalahan yang diambil.

2. Penerapan metode *Naïve Bayes-Support Vector Machine (NB-SVM)* untuk memperoleh model yang digunakan untuk mengetahui dinamika sentimen publik pada masing-masing capres dan cawapres Pemilu 2024.

Pada tahap perancangan model memanfaatkan metode *Naïve Bayes-Support Vector Machine* akan dilakukan beberapa langkah sebagai berikut:

- a. *Data Preparation* yang terdiri atas *Crawling Data*, *Labelling Data*, *Text Pre-processing Data*, *Split Data* dan *Feature Extraction*,
- b. *Sentiment Classifier Modelling*,
- c. *Choose the Best Model (Model Evaluation)*,
- d. *Load the Best Model*,
- e. *Sentiment Prediction* dan
- f. *Result Evaluation*

3. Menyusun Laporan atau Dokumentasi

Tahap akhir adalah penyusunan laporan yang bertujuan untuk mendokumentasikan hasil penelitian serta melampirkan dokumen pendukung lainnya.