

BAB I

PENDAHULUAN

Bab pendahuluan ini membahas mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan yang akan digunakan dalam dokumen skripsi ini.

1.1 Latar Belakang

Pemilihan Kepala Daerah (Pilkada) merupakan proses demokratis yang diadakan untuk memilih kepala daerah, baik gubernur, bupati, maupun walikota, di tingkat provinsi dan kabupaten/kota di Indonesia (Arifulloh, 2016). Pemilihan Kepala Daerah (Pilkada) Jawa Tengah 2024 menjadi sorotan penting dalam konteks politik nasional, mengingat posisi strategis provinsi ini sebagai basis suara terbesar di Indonesia. Dengan pemilihan yang dijadwalkan pada 27 November 2024, kompetisi di Jawa Tengah diperkirakan akan berlangsung sengit, terutama karena tidak adanya petahana yang mencalonkan diri. Hal ini membuka peluang bagi kandidat baru untuk merebut perhatian dan dukungan masyarakat.

Platform X (sebelumnya dikenal sebagai Twitter) memiliki peran yang sangat signifikan dalam media sosial karena memungkinkan komunikasi politik para aktor politik berlangsung secara efektif (Alim & Rahmawati, 2021). Data pada media sosial X, berupa tweet, dapat diolah menjadi informasi yang berguna dan dianalisis. Salah satu analisis yang dapat dilakukan adalah analisis sentimen terhadap opini masyarakat.

Berbagai algoritma *machine learning* telah digunakan pada penelitian mengenai analisis sentimen dengan data yang diambil dari media sosial X seperti penelitian yang dilakukan oleh (Pramesti & Pratiwi, 2023). Pada penelitian tersebut, dilakukan analisis sentimen dengan data *tweet* terkait program MBKM dengan 849 data. Penelitian tersebut membandingkan kinerja *Support Vector Machine* (SVM) dan *Decision Tree* (DT), diperoleh bahwa algoritma *Support Vector Machine* (SVM) memiliki nilai akurasi tertinggi sebesar 85% dibandingkan dengan *Decision Tree* (DT) yang mendapatkan akurasi sebesar 72,86%. Penelitian serupa juga dilakukan oleh Nasution & Hayaty (2019) yang melakukan perbandingan algoritma *Support Vector Machine* (SVM) dengan *K-Nearest Neighbor*

(KNN) pada analisis sentimen dengan data *tweet* yang didapatkan dengan kata kunci “Donald Trump”. Dengan 1.113 data *tweet* diperoleh akurasi sebesar 89,70% pada algoritma SVM yang mengungguli KNN dengan akurasi sebesar 88,10% dengan K=2. Pada penelitian sentimen analisis dari *tweet* di *Twitter* tentang pengaruh pembelajaran daring terhadap motivasi belajar dimasa pandemi COVID-19 yang dilakukan oleh Ariansyah dan Kusmira (2021), penelitian tersebut membandingkan algoritma mana yang memiliki tingkat keakurasian terbaik di antara algoritma *Naïve Bayes* dan *Support Vector Machine*, diperoleh nilai akurasi menggunakan *Support Vector Machine* memiliki akurasi tertinggi sebesar 97,22%, sedangkan nilai akurasi menggunakan algoritma *Naïve Bayes* sebesar 94,79%. Penelitian lain dilakukan Luthfanida (2022) yang melakukan analisis sentimen dari data *Twitter* mengenai wacana tentang “Presiden Jokowi 3 Periode” menggunakan algoritma *Naive Bayes* dan *Support Vector Machine* (SVM), dataset menggunakan data *Twitter* dengan menggunakan kata kunci “Presiden 3 Periode” sebanyak 500 data *tweet*. Berdasarkan berbagai hasil penelitian sebelumnya, algoritma *Support Vector Machine* (SVM) secara konsisten menunjukkan akurasi yang lebih tinggi dibandingkan algoritma lainnya dalam analisis sentimen *tweet* pada berbagai topik. Oleh karena itu, penelitian ini menggunakan SVM sebagai algoritma klasifikasi utama untuk menganalisis sentimen publik mengenai calon gubernur Jawa Tengah periode 2025-2030 pada media sosial X.

SVM mempunyai kelebihan dalam menangani data berdimensi tinggi, tetapi SVM sulit untuk digunakan pada data berskala besar (Nugroho dkk., 2003). Pada umumnya, data *tweet* memiliki atribut atau fitur yang sangat besar, dan jika semua fitur tersebut digunakan, maka akan mengurangi performa algoritma SVM dalam menganalisis sentimen. Untuk menangani kelemahan SVM tersebut, berbagai teknik ekstraksi fitur telah digunakan dalam analisis sentimen untuk merepresentasikan data teks dalam bentuk numerik berupa vektor yang dapat diproses oleh algoritma *machine learning*. Penelitian yang dilakukan oleh Putra dkk. (2023) membandingkan penggunaan teknik ekstraksi fitur TF-IDF dan *Bag of Words* (BoW) pada algoritma SVM untuk analisis sentimen dengan data *tweet* yang diambil dengan pencarian *hashtag* (#jne) sejumlah 300. Penelitian ini menunjukkan bahwa TF-IDF pada SVM mampu menghasilkan akurasi lebih tinggi sedikit dibandingkan *Bag of Words* (BoW) pada SVM dengan akurasi sebesar 86% sedangkan BoW pada SVM sebesar 85%. Penelitian lain oleh Idris dkk. (2025) menunjukkan bahwa *FastText* yang dikembangkan oleh *Facebook Research* menghasilkan akurasi yang sedikit lebih tinggi dibandingkan *Word2Vec* yang

memberikan representasi vektor secara semantik. Penelitian tersebut dilakukan dengan algoritma SVM, *Random Forest*, dan *Gaussian Naïve Bayes* yang dipadukan dengan ekstraksi fitur *Word2Vec* dan *FastText*. Dimana akurasi SVM dengan *FastText* sebesar 89,06%, sedangkan akurasi SVM dengan *Word2Vec* sebesar 89,02%. Dengan demikian, *FastText* terbukti lebih unggul dalam melakukan ekstraksi fitur pada SVM dibandingkan *Word2Vec*.

Berdasarkan masalah, teori, dan fakta yang telah dipaparkan, perlu dilakukan penelitian untuk perbandingan ekstraksi fitur TF-IDF, *Bag of Words*, *Word2Vec*, dan *FastText* pada analisis sentimen publik mengenai calon gubernur Jawa Tengah periode 2025-2030 pada media sosial X menggunakan SVM. Penelitian ini penting untuk memberikan wawasan yang lebih mendalam mengenai efektivitas keempat metode ekstraksi fitur tersebut dalam merepresentasikan teks *tweet* serta menghasilkan akurasi analisis sentimen yang baik dengan algoritma SVM pada berbagai kernel.

1.2 Rumusan Masalah

Pemaparan latar belakang di atas mengarah pada rumusan masalah yang ditetapkan pada penelitian ini, yaitu: “Bagaimana perbandingan akurasi metode ekstraksi fitur TF-IDF, *Bag of Words*, *Word2Vec*, dan *FastText* pada analisis sentimen publik mengenai calon gubernur Jawa Tengah periode 2025-2030 pada media sosial X menggunakan SVM?”

1.3 Tujuan

Permasalahan yang telah dirumuskan mengarah pada tujuan penelitian sebagai berikut:

1. Membandingkan akurasi dari berbagai metode ekstraksi fitur pada analisis sentimen *tweet* publik mengenai calon gubernur Jawa Tengah 2024 dengan algoritma *Support Vector Machine* (SVM).
2. Menganalisis sentimen masyarakat terhadap calon gubernur Jawa Tengah periode 2025-2030 pada media sosial X untuk mengetahui apakah terdapat kesesuaian antara hasil analisis dengan hasil Pilkada Jawa Tengah 2024.

1.4 Manfaat

Penelitian ini diharapkan dapat memberikan wawasan bagi peneliti selanjutnya dalam menentukan teknik ekstraksi fitur yang paling sesuai untuk diterapkan pada algoritma Support Vector Machine (SVM) dalam analisis sentimen berbasis media sosial. Penelitian ini juga dapat menjadi referensi untuk studi selanjutnya mengenai analisis penggunaan media sosial dalam kampanye politik.

1.5 Ruang Lingkup

Berdasarkan rumusan masalah di atas, maka permasalahan yang dibahas dibatasi dengan ruang lingkup berikut:

1. Data yang digunakan dalam penelitian ini diambil dari periode masa kampanye Pilkada Jawa Tengah, yaitu mulai tanggal 25 September 2024 sampai dengan tanggal 23 November 2024.
2. Sumber data yang digunakan pada penelitian ini diperoleh dari metode *crawl data* dari media sosial X dengan menggunakan *library Tweet Harvest*.

1.6 Sistematika Penulisan

Sistematika penulisan memberikan gambaran laporan dari skripsi ini secara urut dan jelas. Berikut adalah beberapa pokok bahasan dalam skripsi ini:

BAB I PENDAHULUAN

Bab ini menyajikan latar belakang masalah, rumusan masalah, tujuan, manfaat, ruang lingkup, serta sistematika penulisan skripsi perbandingan metode ekstraksi fitur TF-IDF, *Bag of Words*, *Word2Vec*, dan *FastText* pada analisis sentimen publik mengenai calon gubernur Jawa Tengah periode 2025-2030 pada media sosial X menggunakan SVM.

BAB II LANDASAN TEORI

Bab ini menyajikan studi pustaka mengenai dasar teori yang digunakan dalam penelitian skripsi ini: *state-of-the-art*, analisis sentimen, pilkada, X, *Natural Language Processing* (NLP), *text mining*, *text preprocessing*, *lexicon based*, *indonesian sentiment lexicon*, *imbalanced data*, *oversampling*, TF-IDF, *Bag*

of Words, *Word2Vec*, *FastText*, *machine learning*, *k-fold cross validation*, *Support Vector Machine (SVM)*, dan evaluasi model.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan metodologi yang digunakan dalam penelitian ini, meliputi alur penelitian, proses pengumpulan data, prapemrosesan data, labelisasi data, pembagian data, TF-IDF, *Bag of Words*, *Word2Vec*, *FastText*, pembangunan model klasifikasi sentimen menggunakan algoritma SVM dengan tiga jenis kernel, yaitu *linear*, RBF, dan *polynomial*. Serta penghitungan klasifikasi dengan data *testing* pada model yang telah dibangun.

BAB IV PEMBAHASAN

Bab ini menjelaskan analisis dan hasil skripsi perbandingan metode ekstraksi fitur TF-IDF, *Bag of Words*, *Word2Vec*, *FastText* pada analisis sentimen publik mengenai calon gubernur Jawa Tengah periode 2025-2030 pada media sosial X menggunakan SVM. Bab ini membahas tentang pelatihan dan skenario pengujian menggunakan tiga jenis kernel SVM: *linear*, RBF, dan *polynomial*.

BAB V PENUTUP

Bab ini berisi kesimpulan dari bab-bab yang telah dijelaskan sebelumnya dan saran untuk pengembangan penelitian lebih lanjut.