

BAB I

PENDAHULUAN

1.1 Latar Belakang

Publik dihebohkan dengan banyak baliho tokoh politik yang terpampang di sudut-sudut kota. Pro kontra muncul dari sejumlah pihak menuding aksi para politikus itu hanya mencari popularitas untuk bekal Pemilihan Presiden (Pilpres) 2024 nanti. Tudingan keras pun sempat muncul saat pemasangan baliho tokoh politik di tengah pandemi Covid-19 dianggap melukai hati nurani masyarakat [1].

Di media sosial twitter Pilpres 2024 juga sedang ramai diperbincangkan, khususnya mengenai baliho sejumlah politisi yang sudah mulai bertebaran bergambar wajah dan logo dari perwakilan beberapa partai yang berukuran super besar di jalanan, walaupun tidak secara eksplisit menunjukkan untuk kepentingan pengenalan diri sebagai calon presiden atau wakil presiden di Pilpres 2024. Namun demikian pertandanya cukup jelas. Meski tak menyatakan langsung mereka hendak maju sebagai calon presiden atau calon wakil presiden, ada tulisan 2024 yang terpampang di baliho. Secara tak langsung, angka 2024 dapat diartikan sebagai Pilpres 2024 [2].

Di era modern ini masyarakat sudah sangat memudahkan dalam menyampaikan sebuah opini terkait suatu hal di sosial media. Aktivitas berinternet yang paling digemari oleh pengguna internet Indonesia ialah bermedia sosial. Saat ini, ada 170 juta jiwa orang Indonesia yang merupakan pengguna aktif media sosial. Rata-rata dari mereka menghabiskan waktu 3 jam 14 menit di platform jejaring sosial [3]. Salah satu hal penting untuk dilihat dari Twitter Page adalah jumlah like atau dapat disebut sebagai dukungan dari para pengguna Twitter

terhadap status ataupun komentar tertentu. Jumlah like dapat diketahui secara otomatis dari Twitter Page, namun dukungan yang berasal dari komentar-komentar yang terdapat pada status Twitter Page tidak bisa dilakukan secara otomatis. Twitter Page belum dapat mengetahui seberapa besar sentimen pengguna (pro atau kontra) dari komentar baik positif ataupun negatif secara otomatis, dengan demikian seberapa sentimen pengguna melalui Twitter Page belum dapat diketahui.

Analisis sentimen adalah suatu bidang yang berlangsung dalam penelitian berbasis teks. Analisis sentimen atau *opinion mining* adalah kajian tentang cara untuk memecahkan masalah dari opini masyarakat, sikap dan emosi suatu entitas tersebut dapat mewakili individu. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung beropini negatif atau positif [4].

Penelitian analisis sentimen telah banyak dilakukan pada media dokumen. Salah satu teknik yang dapat digunakan untuk menganalisis opini ini adalah dengan melakukan klasifikasi pada data mentah dari Twitter. Hal ini dilakukan untuk melihat pendapat atau kecenderungan opini masyarakat apakah cenderung berpandangan positif atau negatif. Ada beragam teknik klasifikasi dokumen, di antaranya adalah *naive bayes classifier*, *decision trees*, dan *support vector machines*.

Naive bayes classifier (NBC) merupakan salah satu algoritma yang digunakan untuk klasifikasi teks serta menggunakan metode Machine Learning dengan perhitungan probabilitas dan statistik yang dikemukakan oleh Thomas Bayes [5]. Pada penelitian ini akan diajukan analisis sentiment terhadap baliho

tokoh politik menjelang Pilpres 2024 melalui dokumen *twitter* dengan menggunakan metode *naive bayes classifier* yang diharapkan mampu menghasilkan nilai akurasi yang tinggi untuk klasifikasi dokumen opini publik sosial media terhadap baliho tokoh politik menjelang Pilpres 2024. Beberapa penelitian telah dilakukan dengan menggunakan metode *naive bayes classifier* diantaranya:

Penelitian sebelumnya berjudul *Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier*. Data pengujian pada penelitian ini sejumlah 796 dokumen tweet dengan 290 sentimen positif dan 506 sentimen negatif. Pengujian dilakukan menggunakan tiga varian jumlah data berbeda, yaitu 100, 200, dan 500 data. Hasil pengujian tertinggi diperoleh dengan menggunakan 100 data dengan accuracy sebesar 67% dan error rate sebesar 33%. Percobaan dengan menggunakan 3 jumlah data berbeda (100, 200, dan 500) menghasilkan selisih nilai accuracy yang tidak jauh berbeda yaitu 0,02. Penelitian ini dilakukan oleh Ni made Ayu Juli Astari, Dewa Gede Hendra Divayana dan Gede Irawan. Penelitian dipublikasikan di *Jurnal Sistem dan Informatika (JSI)*, Vol. 15, No.1, November 2020.

Penelitian lainnya adalah *Naive Bayes Classifier untuk Analisis Sentiment Isu Radikalisme*. Penelitian ini dilakukan oleh Tri A. Sundara, Sherly Eka Putri Arnas dan Sotar pada tahun 2020. Data yang di gunakan dalam penelitian ini, yaitu berupa tweet yang dikumpulkan secara web crawling menggunakan *twitter Application Program Interface (API)*. Data tersebut diolah dengan memanfaatkan *natural language processing (NLP)* dan bahasa pemrograman *python*. Penelitian ini bertujuan untuk menghasilkan informasi sentimen masyarakat terkait

larangan menggunakan cadar dan celana cingkrang bagi Aparatur Sipil Negara. Data yang diperoleh sebanyak 550 data yang dibagi menjadi data latih sebesar 440 dan data uji 110 dengan hasil pengujian akurasi 86%. Penelitian ini dipublikasikan di Prosiding Seminar Nasional Sistem Informasi dan Teknologi (SISFOTEK), Vol. 4, No.1, 19 Agustus 2020 ISSN: 2597-3584.

Penelitian lain dilakukan Arganata Alif Fani dengan judul *Analisis Sentimen Review Pelanggan pada Layanan Indihome Menggunakan Algoritme Naïve Bayes Classifier*. Penelitian ini menampilkan visualisasi dari hasil klasifikasi analisis sentimen dalam bentuk diagram lingkaran (PieChart) dan menghasilkan klasifikasi analisis sentimen yang dibuktikan pada pengujian fungsi Crawling Data, Ekstraksi, Klasifikasi dan Visualisasi dengan nilai sebesar 100%, pengujian perhitungan dengan nilai yang berbeda dengan klasifikasi yang sama, dan pengujian akurasi pada data tahun 2017 sampai dengan 2020 dari jumlah data training sebanyak 40 data dan data testing sebanyak 13 data didapatkan nilai sebesar 69%. Penelitian di muat dalam journal Repositori Universitas Dinamika, 25 Februari 2020.

Penelitian yang dilakukan oleh Yonathan Sari Mahardhika dan Eri Zuliarso (2018) dengan judul *Analisis Sentimen Terhadap Pemerintahan Joko Widodo pada Media Sosial Twitter Menggunakan Algoritme Naïve Bayes Classifier*. Penelitian ini menggunakan data yang berjumlah 400 tweet terdiri dari 300 data latih dan 100 data uji. Data latih merupakan data yang telah diketahui sentimennya, 300 data latih terdiri dari 150 data kelas sentimen negative dan 150 data kelas sentimen positif. Hasil klasifikasi pada data uji menunjukkan, algoritme *naive bayes classifier* memberikan nilai akurasi sebesar 97%. Untuk nilai akurasi tiap

sentimennya yaitu 96% untuk sentimen negative dan 98% untuk sentimen positif. Penelitian di muat dalam Prosiding SINTAK 2018 ISBN: 978-602-8557-20-7.

Penelitian yang dilakukan oleh Martarheza Marthiyas, Budhi Irawan dan Casi Setianingsih (2020) dengan judul *Analisis Sentimen Ulasan Kebijakan Zonasi Sekolah pada Penerimaan Siswa Baru dengan Teks Bahasa Indonesia menggunakan Algoritme Naive Bayes*. Berdasarkan pengujian yang dilakukan, diperoleh persentase 62.39% Negatif, 20.43% Netral dan 17.17% Positif, mayoritas adalah negatif atau tidak menyukai kebijakan tersebut dengan akurasi terbaik pada partisi 90% data latih dan 10% data uji, fitur yang digunakan sebanyak 2000 dengan akurasi sebesar 90.69%, presisi 90.93%, recall 90.69% dan f1 score 90.75% dan waktu pelatihan 0.160 sekon. Penelitian dimuat dalam *Proceeding of Engineering : Vol.7, No.2 Agustus 2020*.

Berdasarkan latar belakang dan permasalahan tersebut, diperlukan penelitian analisis yang bertujuan untuk mengetahui kata-kata yang sering dicuitkan mengenai baliho tokoh politik menjelang Pemilihan Presiden (Pilpres) 2024 dan keterkaitan antar kata, serta mengklasifikasikan opini masyarakat terkait baliho Pemilihan Presiden (Pilpres) 2024 baik secara emosional maupun secara polaritas.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka dirumuskan masalah penelitian ini, yaitu :

1. Bagaimana penerapan algoritme *naïve bayes classifier* terhadap analisis sentiment pada *twitter*?

2. Bagaimana implementasi data uji dan data latih menggunakan algoritme *naïve bayes classifier*
3. Bagaimana hasil sentiment masyarakat Indonesia terhadap baliho tokoh politik menjelang Pilpres 2024?

1.3 Batasan Masalah

1. Dataset yang digunakan khusus *tweet* berbahasa Indonesia dari media sosial *twitter*
2. Kelas sentimen yang digunakan dalam analisis sentimen adalah mencakup opini kelas sentimen positif, negatif dan non opini (netral).
3. Sistem ini menggunakan cuitan pengguna *twitter* sebagai objek penelitian.
4. Dataset yang diambil merupakan *tweet* yang mengandung kata *baliho*, *baliho 2024*, *baliho partai*, *baliho partai politik*, *baliho pilpres*, *baliho pilpres 2024*, *baliho politisi* dan *baliho tokoh politik*.
5. Pengambilan dataset yang digunakan dibatasi dari tanggal 15 Agustus 2021 sampai dengan 25 Agustus 2021
6. kemudian data simulasi dibagi menjadi dua bagian yaitu data training dan data testing.
7. Output dari sistem ini berupa penilaian opini publik dengan melakukan klasifikasi sentimen dan memperhitungkan akurasi atau ketepatan sistem dalam melakukan klasifikasi pada kategori positif, negatif dan non opini (netral).

8. Hasil akhir penelitian yaitu menampilkan data hasil dari analisis sentimen berupa visualisasi, dan wordcloud menggunakan jupyter notebook dengan bahasa pemrograman python.

1.4 Tujuan Penelitian

1. Menerapkan algoritma Naïve Bayes sebagai klasifikasi sentiment masyarakat pada *twitter* terhadap topik baliho tokoh politik menjelang Pilpres 2024.
2. Mengetahui klasifikasi sentimen pada *twitter* terhadap baliho tokoh politik menjelang Pilpres 2024.

1.5 Manfaat Penelitian

1.5.1 Bagi Penulis

1. Penelitian ini diharapkan dapat memberikan gambaran mengenai analisis sentimen pada jejaring sosial twitter dengan mengklasifikasi opini untuk berbagai kepentingan serta pengoptimalan informasi jejaring sosial untuk kepentingan publik.
2. Menerapkan ilmu-ilmu akademis yang didapat selama perkuliahan untuk identifikasi sentimen twitter terkait opini masyarakat terhadap baliho tokoh politik menjelang Pilpres 2024 dengan algoritme *naive bayes classifier*.
3. Menambah pengetahuan dalam bidang data mining dengan metode *naive bayes classifier* khususnya dalam sentimen analisis.

1.5.2 Bagi Universitas

1. Memberikan gambaran terhadap penerapan ilmu pengetahuan yang telah diterima selama kuliah.

2. Menjadi kontribusi literatur karya ilmiah dalam disiplin ilmu teknologi khususnya bidang informatika.

1.6 Metode Penelitian

1.6.1 Metode Pengumpulan Data

Beberapa metode untuk memperoleh data atau informasi dalam menyelesaikan permasalahan adalah sebagai berikut:

1. Studi Kepustakaan

Metode pengumpulan data dengan metode kepustakaan dilakukan dengan pengumpulan jurnal, literatur, paper, makalah, buku, maupun situs internet sebagai sumber pustaka yang berkaitan dengan materi penulisan khususnya analisis sentimen menggunakan algoritma *naive bayes classifier*.

2. Metode Observasi

Observasi dilakukan untuk mengamati objek yang digunakan sebagai sumber data penelitian. Data yang diperoleh merupakan sumber data yang diambil secara langsung dari situs portal *twitter* developer dengan memanfaatkan API (Application Programming Interface) yang disediakan oleh *twitter*. Objek yang diamati adalah tweet dan reply pada *twitter* dengan keyword pencarian baliho dan tweet yang berhubungan dengan baliho tokoh politik menjelang Pilpres 2024. Pada beberapa objek yang diamati terdapat berbagai macam konteksnya, seperti macam-macam opini netizen atau warganet pada *twitter*. Pada tahap ini dilakukan data *crawling* dari *twitter* dengan menggunakan library *python*

tweet. Data yang diambil berupa tweet yang dibatasi dari tanggal 15 Agustus 2021 sampai 25 Agustus 2021.

1.6.2 Metode Analisis

Berikut tahapan metode analisis dalam analisis sentimen :

1. *Data Crawling*

Data tweet didapatkan melalui metode *data crawling* menggunakan library python *tweepy*. Pada tahap ini dilakukan pengambilan data yang tersedia secara online untuk umum dengan cara menginput keyword yang akan dicari, jumlah tweet yang akan diambil, dan tanggal tweet. Tweet yang telah diambil kemudian akan disimpan dalam format csv.

2. *Preprocessing*

Pada preprocessing, dilakukan pembersihan data dari komentar yang telah dikumpulkan dengan melalui beberapa tahapan pemrosesan teks yang terdiri dari *cleansing*, *case folding*, *tokenizing*, *stemming*, *Filtering*. Tahap ini bertujuan supaya data yang didapat akan lebih terstruktur agar lebih mudah untuk dilakukan pengolahan klasifikasi data.

3. Pembobotan

Pada tahap ini dilakukan pembobotan pada tiap dokumen untuk mendapatkan nilai data term/kata yang sudah melalui proses preprocessing. Dengan memakai TF-IDF untuk proses pembobotan. Tahap pembobotan ini dokumen diubah menjadi sebuah vektor dengan

banyak term yang dapat dikenali dari tahapan ekstraksi dokumen menggunakan metode TF-IDF sebagai proses pembobotan.

4. Klasifikasi

Tahap ini adalah proses pengklasifikasian yang akan diproses menggunakan metode Naive Bayes Classifier untuk klasifikasi data mana yang termasuk dalam opini positif, negative maupun non opini.

1.6.3 Metode Perancangan

Penelitian ini menggunakan flowchart untuk menunjukkan tahapan aliran dalam program yang dibuat atau prosedur sistem secara logika dari tahap manipulasi data hingga proses klasifikasi.

1.6.4 Metode Implementasi

Tahap implementasi merupakan penerapan dari metode analisis, seperti *data crawling, preprocessing dan klasifikasi*.

1.6.5 Metode Pengujian

Model yang diperoleh pada proses pelatihan akan dilakukan pengujian *precision, recall, dan f-measure* yang akan disajikan dalam *confussion matrix* untuk mengetahui rata-rata keberhasilan dari model.

1.7 Sistematika Penulisan

Sistematika penulisan skripsi ini terdiri dari lima bab, yang terdiri dari sebagai berikut :

BAB I : PENDAHULUAN

Bab ini membahas mengenai latar belakang masalah, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, metode penelitian, dan sistematika penulisan.

BAB II : LANDASAN TEORI

Bab ini memuat tinjauan pustaka dan dasar – dasar teori yang berkaitan dengan skripsi pada penelitian ini.

BAB III : ANALISIS DAN PERANCANGAN

Bab ini akan membahas mengenai analisis dan model perancangan yang akan digunakan pada sistem.

BAB IV : IMPLEMENTASI DAN PEMBAHASAN

Bab ini menjelaskan tentang implementasi dari perancangan yang telah dibuat pada bab sebelumnya.

BAB V : PENUTUP

Bab ini memuat kesimpulan dari semua hasil penelitian yang telah dilaksanakan dan juga saran – saran yang berkaitan dengan skripsi ini.

