

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Pemilihan Umum adalah proses pemilihan anggota legislatif dan eksekutif yang diselenggarakan serentak di Indonesia pada tanggal 14 Februari 2024 dan dipilih secara langsung. Mekanisme yang diciptakan oleh Badan Pengawas Pemilu (Bawaslu) untuk mewujudkan kedaulatan rakyat dan mewujudkan pemerintahan yang demokratis berdasarkan Pancasila dan Undang-Undang Dasar Negara Republik Indonesia (UUD) Tahun 1945. (Andika dkk., 2019) Menyatakan bahwa “Dalam keberjalanan sistem demokrasi ditandai dengan diadakannya pemilihan umum secara periodik”. Pesta Demokrasi di Indonesia diwujudkan dalam bentuk pemilihan umum presiden dan anggota legislatif yang dilakukan setiap 5 tahun sekali.

Pada hari rabu, tanggal 20 Maret tahun 2024 Komisi Pemilihan Umum telah menetapkan rekapitulasi hasil perolehan suara tingkat Nasional berdasarkan ketentuan Pasal 411 ayat 3 Undang-undang Nomor 7 Tahun 2017 tentang pemilihan umum, Komisi Pemilihan Umum wajib menetapkan secara nasional hasil pemilihan Umum anggota Dewan Perwakilan Rakyat, Dewan Perwakilan Daerah, Presiden dan Wakil Presiden dan hasil pemilihan umum anggota Dewan Perwakilan Rakyat Daerah Provinsi dan Dewan Perwakilan Rakyat Daerah Kabupaten/Kota dengan hasil pemilu presiden dan wakil presiden, pemilu anggota DPR (Dewan Perwakilan Rakyat) di 84 daerah, pemilu anggota DPD (Dewan Perwakilan Daerah) dari setiap Provinsi pada 38 Provinsi, pemilu anggota DPRD (Dewan Perwakilan Rakyat Daerah) Provinsi pada 38 Provinsi di 301 daerah pemilihan umum tahun 2024 dan pemilu anggota DPRD (Dewan Perwakilan Rakyat) Kabupaten/Kota di 2.325 daerah pemilihan dan 508 Kabupaten/Kota tahun 2024 (Komisi Pemilihan Umum, 2024).

Pasangan Calon Presiden dan Wakil Presiden nomor urut 1 H. Anies Baswedan, Ph.D dan Dr. [H.C.] H, A Muhaimin Iskandar dengan perolehan jumlah suara sah sebanyak 40.971.906, Pasangan Calon Presiden dan Wakil Presiden nomor urut 2 H.

Prabowo Subianto dan Gibran Rakabuming Raka dengan perolehan jumlah suara sah sebanyak 96.214.691, Pasangan Calon Presiden dan Wakil Presiden nomor urut 3 H. Ganjar Pranowo, S.H., M.I.P dan Prof. Dr. H. M. Mahfud MD dengan perolehan suara sah sebanyak 27.040.878 (Komisi Pemilihan Umum, 2024). Berdasarkan hasil perolehan tersebut, terjadi perselisihan pendapat di kalangan masyarakat, Selain itu tim yang dipimpin oleh Ganjar-Mahfud menemukan bukti intimidasi dan penghasutan pemilih, hal ini dilakukan untuk memengaruhi keputusan pemilih atau mencegah mereka menggunakan hak mereka untuk memilih. Bawaslu mencatat 2.632 tps yang mengalami kasus intimidasi terhadap pemilih dan penyelenggara pemilihan umum selama pemilu 2024 (TEMPO.CO, 2024). Menyikapi kontroversi yang terjadi, pasangan calon yang berada pada posisi lebih lemah mengajukan hak sengketa kepada Mahkamah Konstitusi. Dengan adanya pengajuan Hak Sengketa hasil pemilihan umum kepada Mahkamah Konstitusi dapat menjadi acuan untuk menyelesaikan problematika perolehan suara dan diharapkan Mahkamah Konstitusi dapat memberikan hasil keputusan sidang sengketa tersebut dengan seadil adilnya tanpa adanya sikap keberpihakan. Hasil dari pemilihan umum tersebut banyak ditanggapi oleh masyarakat melalui media sosial.

Media sosial merupakan alat yang memungkinkan mitra atau pelanggan berkomunikasi dan bertukar informasi menggunakan jaringan. Media sosial mempunyai banyak manfaat, beberapa kegunaan media online antara lain untuk mendapatkan teman baru, mengetahui tentang permainan, aspek keuangan, informasi wisata dan juga isu-isu politik. Saat ini media sosial yang paling populer adalah X, *Facebook*, dan *Instagram* sendiri merupakan aplikasi untuk berbagi foto dan video dengan fungsi seperti *retweet*, mengambil foto dan video serta membagikannya ke berbagai jejaring sosial lainnya. Salah satu topik yang ramai dibicarakan akhir-akhir ini adalah maraknya “calon presiden 2024” di media sosial (Megayasa dkk., 2023). Alasan pemilihan *microblogging X* dikarenakan keterkaitan dengan opini publik dan kemudahan akses dalam pengumpulan data. Dengan mendapatkan akses *API microblogging X* memudahkan untuk mengumpulkan data dengan jumlah yang besar

dan dapat mengatur pencarian berdasarkan kata kunci, lokasi, dan lain sebagainya, sehingga dengan hasil data yang telah diperoleh, dapat dilakukan proses analisis sentimen.

Analisis sentimen merupakan bentuk teknologi baru yang saat ini sedang dikembangkan dan digunakan dalam melakukan sebuah penelitian guna menganalisis topik dan juga pendapat. Teknologi ini sering sekali digunakan untuk melakukan evaluasi dan menganalisis tingkat kepuasan pelanggan atau user terhadap suatu produk atau kebijakan tertentu. Analisis sentimen merupakan proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan sebuah *insight* atau informasi yang berguna (Yahyadi & Latifah, 2022). Dalam proses ini, metode atau algoritma digunakan untuk menentukan apakah teks yang di analisis mengandung sentimen positif, netral atau negatif.

Metode yang digunakan dalam penelitian ini adalah *Naive Bayes Classifier*. Pengklasifikasi *Naive Bayes* adalah metode klasifikasi probabilistik sederhana yang menerapkan teorema independensi tinggi *Bayes*. Algoritma *Naive Bayes* digunakan sebagai pengklasifikasi dalam berbagai masalah dunia nyata, termasuk analisis sentimen, deteksi spam *email*, pengelompokan *email* otomatis, pengurutan prioritas *email*, dan klasifikasi dokumen. Model klasifikasi *Naive Bayes* menghitung probabilitas posterior kelas berdasarkan distribusi kata dalam dokumen. Pengklasifikasi *Naive Bayes* memperkirakan probabilitas setiap kategori dalam fitur dokumen yang diuji, berdasarkan data pelatihan. Sistem dilatih pada data baru (data pelatihan dan data uji) dan diberi tugas untuk menyimpulkan nilai fungsi tujuan dari data tersebut (Andika dkk., 2019).

Dalam melakukan proses analisis sentimen banyak fitur atau kemunculan kata yang mungkin tidak relevan untuk digunakan, oleh karena itu dengan mengurangi jumlah fitur, model hanya dapat menyimpan fitur-fitur penting dan lebih fokus pada teks yang benar-benar penting dalam proses analisis sentimen, sehingga mempercepat proses pelatihan model dan menghemat kebutuhan memori. Mengurangi *noise* atau fitur yang tidak relevan juga dapat membantu meningkatkan akurasi dan performa

model. Penggunaan fitur yang lebih sedikit membuat model lebih cepat dan mudah dilatih untuk melakukan prediksi. Hal ini sangat penting untuk proses analisis yang melibatkan dataset besar atau dalam waktu nyata. *Overfitting* terjadi ketika model terlalu kompleks dan terlalu sesuai dengan data pelatihan, sehingga kurang mampu menangani data baru. Dengan hal ini diharapkan dapat memudahkan pemahaman model dalam membuat keputusan yang penting. Dalam hal ini penulis memilih fitur *Chi-Square* untuk membantu dalam proses analisis sentimen dengan harapan dapat mempermudah penulis dalam menghasilkan analisis sentimen yang akurat. Pemilihan fitur *Chi-Square* bertujuan untuk mengevaluasi pentingnya sebuah kata atau fitur bagi kelas sentimen, kata atau fitur yang digunakan diharapkan dapat mengetahui apakah terdapat hubungan yang signifikan antara dua variabel kategori sehingga, dari hasil tersebut dapat digunakan untuk lebih menyederhanakan kinerja model. Fitur dengan nilai *Chi-Square* yang tinggi dinilai lebih relevan dan informatif karena memiliki hubungan yang erat dengan kelas sentimen. *Chi-Square* merupakan teknik perhitungan yang relatif lebih sederhana dan sangat tepat untuk teks yang mengandung banyak fungsi. *Chi-Square* umumnya lebih mudah digunakan dan seringkali sangat efektif.

Penelitian ini menggabungkan antara metode *Naive Bayes* dan proses seleksi fitur dengan *Chi Square*, Hal ini karena berdasarkan kajian analisis sentimen penerapan PPKM pada *twitter* menggunakan *naive bayes classifier* dengan seleksi fitur *chi square* (Ratiasasadara dkk., 2023) didapatkan hasil penelitian bahwa kemunculan frekuensi kata pada kelas sentiment memiliki peranan penting dalam seleksi fitur *Chi-Square*. Penyeleksian fitur *Chi-Square* menghasilkan 136 dari total 556 fitur atau kata yang memenuhi syarat uji *Chi-Square* dengan mendapatkan tingkat akurasi sebesar 83%. Sehingga penggunaan seleksi fitur *Chi-Square* dapat dikatakan cukup efektif pada model klasifikasi *Naive Bayes*. Data yang digunakan yaitu opini pengguna *X* yang dikelompokkan dengan kelas positif dan negatif. Dan pada penelitian Analisis sentimen menggunakan *Naive Bayes Classifier* dengan *Chi-Square Feature selection* terhadap penyedia layanan telekomunikasi (Nisa dkk., 2019) didapatkan hasil bahwa penerapan seleksi fitur *Chi-Square* dengan tingkat signifikansi 0,01 terhadap klasifikasi

menggunakan metode *naive bayes* memberikan akurasi 85,5%, dan *f1 – score* 84% sedangkan tanpa menggunakan seleksi fitur *chi square* didapatkan akurasi 84,4%, dan *f1 – score* 82%. Penggunaan seleksi fitur *chi square* tidak memberikan pengaruh yang signifikan terhadap hasil klasifikasi *naive bayes*, namun dapat mereduksi fitur yang kurang relevan untuk proses klasifikasi. Data yang digunakan berasal dari opini masyarakat terhadap layanan telekomunikasi. Dan pada penelitian Analisis sentimen terhadap calon wakil presiden Gibran Rakabuming Raka menggunakan algoritma *naive bayes* (Nursyarif dkk., 2024) didapat hasil akurasi sebesar 92,5%, *f1 – score* 92,4%, *Precision* 93,5% dan *Recall* 92,5%. Data yang digunakan yaitu komentar masyarakat di platform youtube. Dapat disimpulkan bahwa dari hasil studi literatur yang telah dilakukan, Penerapan metode *naive bayes* dapat memberikan hasil yang sangat akurat dan penggunaan seleksi fitur *Chi-Square* memungkinkan dapat mereduksi fitur yang tidak penting selama proses klasifikasi.

Pada penelitian ini Analisis sentimen dilakukan dengan tujuan untuk mengetahui sebuah opini masyarakat mengenai Hasil Keputusan Pemilihan Umum tahun 2024 pada *microblogging X* yang dapat digunakan sebagai acuan untuk menentukan apakah opini masyarakat positif, netral atau negatif. Penelitian ini akan menggabungkan metode *Naïve Bayes* dan penyeleksian fitur dengan menggunakan *Chi-Square*, sehingga nantinya akan menghasilkan dua perbandingan hasil klasifikasi dengan menggunakan *chi-square* dan tanpa menggunakan *chi-square*. Tahapan *Chi Square* (Seleksi fitur) bertujuan untuk menghapus kata yang tidak berdampak signifikan pada hasil sentimen dari sejumlah data yang telah dilakukan *pre-processing*, sehingga fitur atau kata yang dihasilkan diharapkan dapat meningkatkan hasil akurasi dalam proses pengklasifikasian agar lebih efektif. Berdasarkan data yang dikumpulkan selama proses analisis sentimen, penulis berharap dapat mengetahui opini masyarakat apakah lebih cenderung berpendapat positif, netral, atau negatif. Dengan demikian, penelitian ini diharapkan dapat menjadi bahan pertimbangan bagi pemerintah untuk

membantu mengevaluasi penyelenggaraan pemilihan umum pada periode berikutnya agar lebih baik.

1.2 Rumusan Masalah

Berdasarkan penjelasan Latar Belakang diatas maka didapat rumusan masalah dalam penelitian ini, yaitu:

- a. Bagaimana cara melakukan proses crawling data untuk mengetahui opini masyarakat terkait Hasil Keputusan Pemilu Tahun 2024 dengan metode *Naïve Bayes*?
- b. Bagaimana cara melakukan proses analisis data Hasil Keputusan Pemilu Tahun 2024 dengan metode *Naïve Bayes*?
- c. Bagaimana klasifikasi sentimen Hasil Keputusan Pemilu Tahun 2024 dengan menggunakan *Naive Bayes* dan pengaruh penggunaan seleksi fitur *Chi-square*?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

- a. Untuk mengetahui opini masyarakat terkait Hasil Keputusan Pemilihan umum Tahun 2024 dengan metode *Naive Bayes*.
- b. Dapat mengetahui hasil akurasi analisis data Hasil Keputusan Pemilihan umum Tahun 2024 dengan metode *Naive Bayes*.
- c. Dapat mengetahui pengaruh penggunaan seleksi fitur *chi-square* terhadap proses klasifikasi dengan *Naïve Bayes*.

1.4 Manfaat

Manfaat dari penelitian ini adalah:

- a. Menambah wawasan mengenai proses analisis data pada *microblogging X*.
- b. Bermanfaat untuk menambah informasi terkait pengaruh penggunaan seleksi fitur *chi-square* terhadap klasifikasi dengan metode *naïve bayes* sehingga diharapkan menghasilkan ketepatan algoritma dan menghasilkan fitur yang memiliki hubungan signifikan antara variabel kategori.
- c. Hasil dari penelitian ini dapat dijadikan sumber acuan untuk penelitian kedepannya mengenai analisis sentimen, dan juga dapat dijadikan sebagai

bahan pertimbangan bagi pemerintah untuk dapat mengevaluasi pelaksanaan pemilihan umum pada periode selanjutnya agar menjadi lebih baik.