

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dewan Perwakilan Rakyat (DPR) merupakan lembaga legislatif yang berperan sebagai wakil rakyat dan memiliki peranan strategis dalam menjalankan tiga fungsi fundamental, yaitu pengambilan keputusan, pembentukan kebijakan, dan pengawasan terhadap pemerintah. DPR memengaruhi berbagai aspek kehidupan masyarakat, seperti ekonomi, sosial, dan kebijakan publik yang diharapkan dapat mewujudkan pemerintahan yang demokratis [1]. Dalam beberapa tahun terakhir, DPR telah menimbulkan berbagai kontroversi yang memicu sentimen negatif dari masyarakat, salah satunya terkait isu kenaikan gaji anggota DPR. Sentimen negatif tersebut semakin kuat karena kenaikan gaji sering kali terjadi ketika pemerintah menyerukan penghematan atau saat harga kebutuhan pokok mengalami kenaikan, sehingga menimbulkan persepsi bahwa para pejabat negara kurang peka terhadap kondisi masyarakat [2]. Hal itu membuat kepercayaan masyarakat kepada anggota DPR kian tergerus [3]. Masyarakat luas juga mengekspresikan kritik terhadap DPR yang dianggap tidak mengutamakan kepentingan rakyat [4]. Akibatnya, ketidakpedulian masyarakat terhadap politik meningkat, partisipasi pemilihan umum menurun, dan peningkatan aksi protes yang dapat mengganggu keamanan sosial [3]. Pada era digital ini, diskusi politik banyak dilakukan di media sosial, salah satunya adalah TikTok [5]. Di platform ini, masyarakat mengekspresikan beragam persepsi mereka tentang kebijakan DPR [6]. Komentar TikTok yang sangat banyak dan tidak terstruktur serta beragam bahasa dan gaya bahasa membuat pengkajian pendapat publik sulit dilakukan melalui pendekatan manual [7]. Oleh karena itu, penting dilakukan analisis sentimen pengguna TikTok terhadap isu kenaikan gaji anggota DPR untuk menghasilkan data terstruktur mengenai kecenderungan opini masyarakat yang diklasifikasikan ke dalam sentimen positif, negatif, dan netral. Hasil analisis tersebut diharapkan dapat menjadi bahan evaluasi serta acuan bagi DPR dalam memahami kecenderungan opini dan tingkat kepercayaan masyarakat terhadap lembaga legislatif.

Penelitian terkait analisis sentimen sebelumnya telah dilakukan menggunakan *web scraping* dan *Naïve Bayes Classifier* untuk analisis sentimen berita politik di YouTube. Dalam penelitian tersebut, diterapkan kerangka kerja KDD yang mencakup lima langkah

berurutan: memilih data yang relevan, membersihkannya melalui proses pra-pemrosesan, mengubahnya ke dalam format yang sesuai, melakukan eksplorasi data, dan mengevaluasi temuan yang diperoleh. Data diambil dari 1.000 komentar pada lima video berita politik di kanal YouTube Detik.com, yang diklasifikasikan menjadi sentimen positif dan negatif. Hasil evaluasi menunjukkan bahwa klasifikasi menggunakan *Naïve Bayes* menghasilkan tingkat akurasi sebesar 87% pada skema validasi dengan pembagian 90% data latih dan 10% data uji. Namun, klasifikasi sentimen positif memiliki nilai *F1-score* lebih rendah, menunjukkan kecenderungan model dalam mengenali sentimen negatif. Penelitian ini dilakukan untuk mengatasi beberapa kesenjangan pada penelitian sebelumnya yang diuraikan dalam bagian kesimpulan dan saran, yaitu perlunya menyeimbangkan dataset, pengujian algoritma selain *Naïve Bayes* untuk menentukan algoritma paling optimal dalam analisis sentimen pada berita politik, penambahan jumlah data, pertimbangan klasifikasi sentimen menjadi tiga kategori (positif, negatif, netral), serta penggunaan leksikon sentimen yang lebih komprehensif untuk pelabelan data [8].

Dalam penelitian ini, proses yang dilakukan meliputi pengumpulan dan pemisahan dataset secara manual menjadi data latih dan data uji. Data latih akan mengalami proses pra-pemrosesan teks (*data cleaning, normalization, stopword removal, tokenization, dan stemming*), pelabelan data dengan metode *lexicon-based* menggunakan kamus InSet, ekstraksi fitur menggunakan TF-IDF, serta penyeimbangan kelas dengan SMOTE. Selanjutnya, dilakukan pelatihan model *Naïve Bayes* dan SVM. *Naïve Bayes* dipilih karena memiliki proses komputasi yang sederhana dan efisien untuk klasifikasi teks, sedangkan SVM dipilih karena dikenal memiliki kemampuan yang baik dalam mengklasifikasikan data berdimensi tinggi, seperti hasil ekstraksi fitur TF-IDF pada analisis sentimen, hingga diperoleh model terlatih. Sementara itu, data uji menjalani tahapan pra-pemrosesan, pelabelan, dan ekstraksi fitur yang serupa, kemudian diklasifikasikan oleh kedua model. Hasil klasifikasi tersebut selanjutnya dievaluasi menggunakan metrik *accuracy, precision, recall, dan F1-score* untuk membandingkan kinerja *Naïve Bayes* dan SVM. Berdasarkan penjelasan yang telah disampaikan, penelitian ini akan disusun dalam bentuk tugas akhir dengan judul "ANALISIS SENTIMEN OPINI PUBLIK TERHADAP ISU KENAIKAN GAJI ANGGOTA DPR PADA PLATFORM TIKTOK MENGGUNAKAN NAIVE BAYES DAN SUPPORT VECTOR MACHINE".

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang yang telah dipaparkan, permasalahan dalam penelitian ini adalah banyaknya komentar pengguna TikTok terkait isu kenaikan gaji anggota DPR yang tidak terstruktur menyebabkan pengkajian pendapat sulit dianalisis secara manual. Oleh karena itu, diperlukan analisis sentimen untuk mengklasifikasikan komentar ke dalam kategori positif, negatif, dan netral menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM).

1.3 Tujuan

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Membangun aplikasi analisis sentimen untuk mengklasifikasikan komentar pengguna TikTok terkait isu kenaikan gaji anggota DPR ke dalam kategori positif, negatif, dan netral menggunakan metode *Naïve Bayes* dan *Support Vector Machine* (SVM).
2. Melakukan evaluasi terhadap hasil klasifikasi sentimen yang dihasilkan oleh algoritma *Naïve Bayes* dan *Support Vector Machine* (SVM) berdasarkan nilai *accuracy*, *precision*, *recall*, dan *F1-score*.

1.4 Manfaat

Adapun manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Memudahkan proses pengklasifikasian komentar masyarakat terhadap isu kenaikan gaji anggota DPR pada platform TikTok.
2. Hasil klasifikasi sentimen positif, negatif, dan netral dapat dimanfaatkan sebagai bahan evaluasi serta menjadi acuan bagi DPR dalam memahami tingkat kepercayaan masyarakat.

1.5 Ruang Lingkup

Untuk memastikan pembahasan dalam tugas akhir ini tetap fokus dan terarah, diperlukan penetapan ruang lingkup yang jelas sebagai berikut:

1. Seluruh data diperoleh melalui platform Apify dengan metode *web scraping* dalam format .xlsx sebanyak 7.000 komentar berbahasa Indonesia dari pengguna TikTok dalam rentang waktu April 2025 hingga Januari 2026.
2. Data diambil menggunakan kata kunci "Kenaikan gaji anggota DPR".
3. Dataset dibagi menjadi rasio 80:20 secara manual, dengan 80% digunakan sebagai data latih dan 20% sebagai data uji.

4. Dataset akan diproses melalui tahap pra-pemrosesan teks yang meliputi *data cleaning*, *normalization*, *stopword removal*, *tokenization*, dan *stemming*.
5. Proses pelabelan data dilakukan secara otomatis menggunakan metode *lexicon-based* dengan kamus InSet tanpa dilakukan pemeriksaan ulang secara manual.
6. Ekstraksi fitur dilakukan menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF).
7. Kelas minoritas pada data latih diseimbangkan menggunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE).
8. *Naïve Bayes* dilatih menggunakan parameter *default* ($\alpha = 1,0$) dan tidak dilakukan proses *tuning parameter* maupun pencarian nilai *alpha* terbaik.
9. SVM menggunakan kernel linear dengan parameter lainnya mengikuti nilai *default Scikit-Learn* dan tidak dilakukan iterasi pencarian parameter maupun *tuning parameter*.
10. Kinerja model klasifikasi dievaluasi menggunakan *confusion matrix* untuk memperoleh nilai *accuracy*, *precision*, *recall*, dan *F1-score*.
11. Komentar yang dianalisis dalam penelitian ini dikategorikan ke dalam tiga sentimen, yaitu positif, negatif, dan netral.