

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam era digital saat ini, internet menjadi aspek kehidupan yang tidak bisa terpisahkan dengan kehidupan masyarakat Indonesia. Perkembangan teknologi informasi telah membawa perubahan besar dalam cara masyarakat menyampaikan opini, terutama melalui media sosial. Media sosial berperan penting dalam menggali informasi termasuk berita dalam negeri maupun luar negeri. Media sosial menjadi wadah utama sekaligus jembatan komunikasi bagi masyarakat menyampaikan aspirasi karena terbatasnya ruang berbicara secara langsung (Apriyani, 2021).

Media sosial X (sebelumnya Twitter) menjadi salah satu *platform* yang banyak digunakan masyarakat Indonesia untuk mengekspresikan pandangan terhadap isu-isu politik, sosial, maupun kebijakan pemerintah (Fatkhudin dkk., 2024). Masyarakat memanfaatkan media sosial sebagai media utama dalam merespon kebijakan pemerintah melalui unggahan teks, komentar, dan penggunaan tagar yang merepresentasikan sikap emosional terhadap keputusan pemerintah (Lestari & Berliani, 2024). Salah satu isu yang menimbulkan kontroversi adalah kebijakan kenaikan tunjangan anggota DPR yang menimbulkan diskursus publik luas. Isu ini ramai diperbincangkan dan memunculkan beragam opini pro dan kontra. Respon masyarakat tercermin melalui petisi daring, aksi demonstrasi Mahasiswa, serta percakapan publik dalam komentar konten media sosial yang tersebar dalam jagat maya sehingga menimbulkan keributan dalam aksi demonstrasi yang merugikan pribadi maupun negara (Pratama dkk., 2025).

Menurut Cahyaningtyas dkk (2021), analisis sentimen diperlukan untuk mengetahui kecenderungan opini masyarakat. Namun, dalam pengolahan data teks, sering ditemui masalah berupa jumlah data yang tidak seimbang (*imbalanced dataset*), dimana jumlah data sentimen positif atau negatif tidak merata (Rahma & Suadaa, 2023). Hal ini dapat menurunkan kinerja algoritma klasifikasi dalam menghasilkan prediksi yang akurat. Penelitian sebelumnya telah banyak membahas analisis sentimen pada isu sosial dan politik menggunakan berbagai algoritma

klasifikasi. Misalnya, penelitian mengenai *analisis sentimen Pemilu 2024 di media sosial X* yang dilakukan oleh Khoirunnisa dkk (2024), menunjukkan bahwa kelas sentimen positif mendominasi dataset dengan hasil akurasi KNN meraih 97,50%, *Decision Tree* 97,25%, dan *Naïve Bayes* 96,14%, menunjukkan *Decision Tree* tetap kompetitif, sehingga menyebabkan ketidakseimbangan data yang berdampak pada bias model dan akurasi yang menurun. Meski algoritma seperti *Decision Tree*, *Naïve Bayes*, dan KNN menghasilkan akurasi tinggi, model tetap cenderung mengabaikan kelas minoritas. Penelitian tersebut belum menerapkan strategi *text augmentation* sebagai pendekatan untuk menyeimbangkan data sentimen yang tidak seimbang, sehingga potensi peningkatan performa model, terutama *Decision Tree*, belum dieksplorasi secara optimal.

Berdasarkan penelitian Aby Vonega dkk (2022) mengenai Analisis Sentimen Twitter terhadap Opini Publik atas Isu Pencalonan Puan Maharani dalam PILPRES 2024, penelitian ini bertujuan mengeksplorasi opini publik dengan metode *text mining* menggunakan *library TextBlob*, *VaderSentiment*, dan *SentiWordNet*. Hasil penelitian menunjukkan bahwa opini publik di twitter didominasi oleh sentimen netral sebesar 97,25%, diikuti sentimen positif sebesar 2,55%, dan negatif hanya 0,1%, sehingga dapat disimpulkan bahwa masyarakat cenderung bersikap netral terhadap isu pencalonan tersebut. Penelitian ini belum membahas solusi terhadap dominasi sentimen netral yang menimbulkan ketidakseimbangan data dan tidak menguji peningkatan performa model melalui *text augmentation* dalam konteks isu kebijakan publik.

Penelitian lainnya dilakukan oleh Arham dkk (Arham dkk., 2022), mengenai analisis sentimen pada opini masyarakat terhadap Virus Covid-19 Varian *Omicron* ini membandingkan kinerja tiga algoritma klasifikasi dalam menganalisis 3.931 data *tweet* yaitu Algoritma *Naïve Bayes*, *Decision Tree*, dan *Support Vector Machine*. Hasil penelitian menunjukkan bahwa algoritma *Support Vector Machine* (SVM) memberikan akurasi tertinggi sebesar 81,68%, sedangkan *Naïve Bayes* memperoleh akurasi 74,25% dan *Decision Tree* sebesar 72,77%. Analisis sentimen juga memperlihatkan bahwa opini masyarakat didominasi oleh sentimen netral (63,7%), sementara sentimen negatif 21,1% dan positif 15,2%, yang mengindikasikan bahwa masyarakat lebih banyak menanggapi varian *Omicron*

secara netral. Fokus penelitian masih pada perbandingan algoritma tanpa mempertimbangkan optimalisasi dataset menggunakan *text augmentation*, padahal pendekatan ini berpotensi meningkatkan kinerja model sederhana seperti *Decision Tree* pada data sentimen yang tidak seimbang.

Permasalahan *imbalanced data* sering ditemui pada klasifikasi teks, di mana model cenderung hanya berfokus pada kelas mayoritas dan mengabaikan kelas minoritas. Kondisi ini dapat menurunkan performa model, terutama ketika akurasi prediksi pada kelas minoritas memiliki tingkat kepentingan yang lebih tinggi. Salah satu pendekatan yang dapat digunakan untuk mengatasinya adalah *text augmentation*, seperti *synonym replacement* dan *back translation* (Rahma & Suadaa, 2023). *Text augmentation* atau augmentasi data teks merupakan sekumpulan teknik yang digunakan untuk menghasilkan data teks sintetis dari data asli yang telah tersedia (Harywanto dkk., 2022). Berbagai penelitian sebelumnya telah banyak menerapkan teknik *text augmentation* pada berbagai domain analisis sentimen, namun penerapannya pada topik kebijakan publik di media sosial, khususnya terkait kebijakan kenaikan tunjangan DPR, masih sangat terbatas. Oleh karena itu, penelitian ini berfokus pada penerapan metode *synonym replacement* dan *back translation* untuk meningkatkan kualitas data serta performa model klasifikasi menggunakan algoritma *Decision Tree CART* pada analisis sentimen kebijakan kenaikan tunjangan DPR di media sosial X (Rahma & Suadaa, 2023).

Teknik *text augmentation* telah berkembang pesat sebagai metode efektif untuk meningkatkan kualitas dan jumlah data dalam tugas analisis sentimen. Dengan memodifikasi data teks asli melalui berbagai cara transformasi, teknik ini mampu mengatasi masalah ketidakseimbangan data (*imbalanced dataset*) yang sering ditemukan dalam pemrosesan bahasa alami (Rahma & Suadaa, 2023). Pendekatan terbaru seperti *Text Augmentation with Word Contributions* (TAWC) yang dikembangkan oleh Santoso dkk (Santoso dkk., 2024), menunjukkan peningkatan signifikan pada performa model klasifikasi sentimen, dengan kemampuan mempertahankan makna semantik dan konteks kata dalam kalimat. Meskipun menunjukkan hasil menjanjikan, pendekatan ini belum diuji secara luas pada dataset berbahasa Indonesia dengan *domain* kebijakan publik, khususnya dalam konteks analisis sentimen terhadap isu politik seperti kenaikan tunjangan DPR.

Analisis sentimen terhadap opini publik di media sosial memiliki peranan penting dalam mendukung pembuatan kebijakan publik yang responsif dan tepat sasaran. Dalam era digital, data yang berasal dari media sosial seperti twitter dapat memperlihatkan tren sentimen masyarakat secara *real-time*, memberikan gambaran yang relevan dan berkelanjutan terhadap persepsi publik terkait kebijakan pemerintah (Ceron & Negri, 2021). Sentimen positif maupun negatif yang muncul di ruang digital ini tidak hanya memengaruhi citra sosial-politik, namun juga dapat menunjang proses evaluasi dan penyesuaian kebijakan secara proaktif oleh pengambil keputusan. Oleh karena itu, penelitian ini menempatkan analisis sentimen sebagai alat strategis untuk mengevaluasi kebijakan kenaikan tunjangan DPR melalui data media sosial, memberikan *insight* yang bermanfaat bagi transparansi dan akuntabilitas publik (Pratama dkk., 2025).

Studi yang dilakukan oleh Hardyatman dan Hasan (2025), mengenai analisis sentimen terkait kebijakan perpajakan bertujuan untuk menganalisis sentimen masyarakat terhadap rencana kenaikan PPN 12% di Indonesia menggunakan *Decision Tree* menunjukkan bahwa *Decision Tree* berhasil meningkatkan akurasi sebesar 81,34% klasifikasi sentimen dibandingkan metode sebelumnya, seperti *Naïve Bayes* dengan tingkat akurasi 63,1% sehingga terbukti mampu memberikan hasil klasifikasi yang cukup baik meskipun menggunakan data yang kompleks. Meskipun *Decision Tree* terbukti efektif untuk data kompleks, belum ada penelitian yang mengombinasikan algoritma ini dengan teknik *text augmentation* untuk mengatasi *imbalanced dataset* pada analisis sentimen kebijakan publik di media sosial.

Penerapan teknik *text augmentation* untuk analisis sentimen kebijakan kenaikan tunjangan DPR perlu mempertimbangkan berbagai aspek, antara lain pemilihan teknik augmentasi yang tepat, penanganan ketidakseimbangan kelas sentimen, optimasi parameter model, serta evaluasi performa secara menyeluruh menggunakan metrik yang relevan. Penerapan *text augmentation* dilakukan menggunakan tiga teknik yaitu *Synonym Replacement*, *Back Translation*, dan kombinasi keduanya, yang diterapkan pada data komentar media sosial X untuk menyeimbangkan distribusi kelas sentimen sebelum diklasifikasikan menggunakan algoritma *Decision Tree* CART yang dioptimasi dengan *Grid Search Cross*

Validation. *Text augmentation* terbukti mampu mengatasi ketidakseimbangan data, meningkatkan kemampuan model dalam mengenali kelas minoritas, serta menghasilkan klasifikasi sentimen yang lebih akurat dan seimbang antara kelas positif dan negatif. Penelitian ini juga dapat membuktikan secara empiris bahwa teknik *text augmentation* berbasis *Synonym Replacement* dan *Back Translation* efektif meningkatkan performa model *Decision Tree* CART pada data sentimen berbahasa Indonesia yang tidak seimbang. Hasil penelitian ini dapat memberikan kontribusi nyata bagi pengembangan analisis sentimen berbahasa Indonesia, khususnya sebagai acuan dalam penerapan teknik augmentasi teks untuk menangani ketidakseimbangan data pada klasifikasi sentimen isu kebijakan publik.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana pengaruh penerapan teknik *text augmentation* (*Synonym Replacement*, *Back Translation*, dan kombinasi keduanya) terhadap performa model *Decision Tree* CART dalam mengklasifikasikan sentimen publik terkait kebijakan kenaikan tunjangan DPR di media sosial X?
2. Teknik *text augmentation* manakah yang menghasilkan performa model *Decision Tree* CART paling optimal berdasarkan nilai akurasi dan f1-score?
3. Bagaimana perbandingan performa model *Decision Tree* CART sebelum dan sesudah penerapan *text augmentation* dalam mengatasi ketidakseimbangan kelas sentimen pada data komentar media sosial X?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Menganalisis pengaruh penerapan teknik *text augmentation* (*Synonym Replacement*, *Back Translation*, dan kombinasi keduanya) terhadap performa model *Decision Tree* CART dalam mengklasifikasikan sentimen publik terkait kebijakan kenaikan tunjangan DPR di media sosial X.

2. Menentukan teknik *text augmentation* yang menghasilkan performa model *Decision Tree* CART paling optimal berdasarkan nilai akurasi dan f1-score pada data komentar media sosial X.
3. Membandingkan performa model *Decision Tree* CART sebelum dan sesudah penerapan *text augmentation* dalam menangani ketidakseimbangan kelas sentimen pada data komentar media sosial X terkait kebijakan kenaikan tunjangan DPR.

1.3.2 Manfaat Penelitian

Penelitian ini memberikan kontribusi ilmiah dalam pengembangan analisis sentimen berbahasa Indonesia menggunakan algoritma *Decision Tree* dan teknik *text augmentation*. Kajian ini dapat menjadi rujukan akademik dalam penerapan optimasi parameter menggunakan *Grid Search Cross Validation* untuk meningkatkan performa model klasifikasi teks. Penelitian juga menyediakan gambaran empiris mengenai persepsi masyarakat terhadap kebijakan kenaikan tunjangan DPR yang dapat dijadikan bahan pertimbangan bagi pihak pembuat kebijakan dalam menilai respons publik. Dari sisi teknis, penelitian ini memberikan alternatif penerapan teknik augmentasi sederhana seperti *Synonym Replacement*, *Back Translation*, dan kombinasi keduanya sebagai solusi efisien untuk penyeimbangan data. Pendekatan ini dapat diterapkan pada sistem analisis sentimen di bidang lain dengan karakteristik data yang serupa.

1.4 Batasan Masalah

Penelitian ini menetapkan batasan masalah untuk memastikan fokus dan kejelasan dalam pembahasan yang dilakukan. Adapun batasan masalah dari penelitian ini adalah sebagai berikut:

1. Data yang digunakan berupa komentar dari Twitter dengan kata kunci terkait kebijakan kenaikan tunjangan DPR.
2. Sentimen dibatasi pada dua kategori: positif dan negatif.
3. Teknik *text augmentation* yang digunakan adalah *Synonym Replacement*, *Back Translation*, dan kombinasi keduanya.
4. Representasi teks menggunakan TF-IDF, sedangkan algoritma klasifikasi menggunakan *Decision Tree*.

5. Algoritma klasifikasi yang digunakan adalah *Decision Tree* dengan optimasi parameter melalui *Grid Search Cross Validation*.
6. Pada penelitian ini untuk analisis dan pemrosesan data menggunakan *Python* dan *Google Collab*.

1.5 Sistematika Penulisan Proposal Skripsi

Sistematika penulisan skripsi ini disusun dengan mengikuti aturan penulisan skripsi pada program studi Sistem Informasi Universitas Trunodjoyo Madura. Berikut metodologi penulisan skripsi pada penelitian ini.

BAB I PENDAHULUAN

Bab ini menjelaskan mengenai latar belakang, rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, serta sistematika penulisan terkait penelitian yang dilakukan.

BAB II TINJAUAN PUSTAKA

Bab ini membahas mengenai landasan teori terkait kebijakan kenaikan tunjangan DPR serta metode algoritma *Decision Tree*. Bab ini juga merinci proses penelitian, mulai dari teknik *pre-processing*, teknik augmentasi, pelabelan, ekstraksi fitur hingga evaluasi model, serta mencakup penelitian terdahulu yang relevan.

BAB III METODE PENELITIAN

Bab ini menjelaskan mengenai metode penelitian yang dilakukan, yaitu mengenai dataset yang digunakan, arsitektur sistem, serta skenario-skenario yang di uji coba.

BAB IV HASIL DAN PEMBAHASAN

Bab ini membahas hasil implementasi metode *Decision Tree* untuk klasifikasi sentimen pada data isu terkait kebijakan kenaikan tunjangan DPR. Pengujian dilakukan dengan teknik penyeimbangan data *text augmentation*. Evaluasi dilakukan berdasarkan akurasi, presisi, *recall*, dan *f1-score*, serta analisis pengaruh kernel dan parameter terhadap performa model.

BAB V KESIMPULAN DAN SARAN

Bab ini membahas kesimpulan dari hasil penelitian yang didapatkan selama melakukan penelitian beserta saran-saran untuk penelitian selanjutnya.