

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi telah menjadikan media sosial sebagai ruang publik digital utama bagi pertukaran opini di masyarakat. Salah satu platform yang menempati posisi strategis karena sifatnya yang *real-time* dan kemampuannya memfasilitasi penyebaran informasi serta pembentukan opini secara masif di Indonesia adalah X, yang sebelumnya dikenal sebagai Twitter (Faidh et al., 2024). Melalui platform ini, masyarakat dapat menyuarakan kritik, dukungan, maupun pandangan terhadap berbagai isu dan kebijakan yang tengah berkembang, sehingga tercipta rekam jejak digital yang kaya akan data tekstual dan berpotensi dimanfaatkan untuk mengkaji pola pikir serta sentimen publik.

Dalam beberapa tahun terakhir, isu-isu strategis seperti kebijakan fiskal, program bantuan sosial, dan regulasi pemerintah banyak menjadi bahan diskusi warganet melalui cuitan yang bersifat spontan dan emosional, yang kerap merepresentasikan sikap nyata terhadap kebijakan pemerintah (Pramudita et al., 2024). Di tengah dinamika tersebut, Rancangan Undang-Undang (RUU) Perampasan Aset muncul sebagai salah satu polemik hukum yang menyita perhatian publik karena dirancang untuk memperkuat penelusuran, pembekuan, dan perampasan aset yang diduga berasal dari tindak pidana, khususnya korupsi dan pencucian uang, guna meningkatkan efektivitas pemulihan kerugian negara (Nugroho et al., 2023). Sebagian kalangan memandang RUU ini sebagai instrumen penting untuk memperkuat *asset recovery* dan menutup celah bagi pelaku kejahatan ekonomi, sementara pihak lain mengkhawatirkan potensi penyalahgunaan kewenangan serta risiko terhadap perlindungan hak-hak warga negara, transparansi, dan akuntabilitas pelaksanaannya (Rofiqi & Akbar, 2024).

Perbedaan pandangan tersebut tercermin dalam beragam opini di media sosial, termasuk X, di mana warganet menyuarakan dukungan maupun penolakan terhadap RUU Perampasan Aset. Kondisi ini menegaskan pentingnya pemetaan sentimen publik yang dilakukan secara sistematis agar persepsi masyarakat terhadap RUU tersebut dapat dipahami secara lebih komprehensif dan terukur. Analisis sentimen berbasis *Natural Language Processing* (NLP) memungkinkan opini kualitatif yang

tersebar di media sosial dikonversi menjadi informasi terstruktur berupa kategori sentimen, sehingga pemangku kebijakan dapat memperoleh gambaran yang lebih objektif mengenai tingkat penerimaan publik terhadap suatu regulasi (Pramudita et al., 2024).

Selain analisis sentimen, pemahaman terhadap isi atau tema utama dari opini publik juga menjadi aspek penting. Salah satu pendekatan yang dapat digunakan untuk mengekstraksi tema tersembunyi dari data teks adalah metode *Latent Dirichlet Allocation* (LDA). Metode ini memungkinkan setiap dokumen direpresentasikan sebagai distribusi probabilitas terhadap sejumlah topik, sehingga dapat mengungkap pola pembahasan utama yang muncul dalam kumpulan cuitan sehingga mampu memahami konteks isu yang mendasari dukungan maupun penolakan terhadap suatu kebijakan.

Berdasarkan kompleksitas opini publik yang berkembang, studi ini secara spesifik meneliti penerapan metodologi CRISP-DM dan algoritma *deep learning* sebagai pendekatan yang efektif untuk memetakan sentimen masyarakat terhadap RUU Perampasan Aset. Metode CRISP-DM (*Cross Industry Standard Process for Data Mining*) yaitu salah satu metodologi yang dapat digunakan selama proses analisis data tekstual media sosial. CRISP-DM mengadopsi sistem yang teratur yang terdiri dari enam tahap utama: pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penerapan. Dengan mengacu pada kerangka ini, proses penelitian dapat dilakukan secara runtut, mulai dari merumuskan tujuan analisis sentimen terhadap RUU Perampasan Aset, mengumpulkan dan menyiapkan data cuitan, membangun model klasifikasi, hingga mengevaluasi kinerja model yang dihasilkan (Amalsyah et al., 2025).

Pada konteks media sosial X, tahapan tersebut diawali dengan pengumpulan data melalui proses *crawling* menggunakan kata kunci yang relevan dengan RUU Perampasan Aset, sehingga hanya cuitan yang terkait topik tersebut yang dihimpun (Ibrahim et al., 2025). Data mentah kemudian diproses melalui serangkaian tahapan pra-pemrosesan teks, seperti *cleansing*, *case folding*, normalisasi kata tidak baku, penghapusan tautan dan simbol tidak informatif, serta tokenisasi, sebelum selanjutnya diberi label sentimen positif atau negatif menggunakan pendekatan berbasis leksikon melalui *SentiWordNet*. *SentiWordNet* digunakan untuk

memberikan skor polaritas terhadap kata dalam teks, kemudian skor tersebut diakumulasi untuk menentukan kecenderungan sentimen pada setiap cuitan. Setelah pelabelan, teks diubah menjadi representasi numerik menggunakan teknik ekstraksi fitur yang mampu menangkap hubungan semantik antar kata, dan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) dapat diterapkan untuk menyeimbangkan distribusi kelas pada data latih jika terjadi ketidakseimbangan sentiment (Ibrahim et al., 2025) (Verdikha & Yulianto, 2025). Data yang telah siap kemudian diproses menggunakan algoritma *Bidirectional Long Short-Term Memory* (BiLSTM) yang unggul dalam mempelajari konteks kalimat secara dua arah. Tahap-tahap tersebut disusun mengikuti alur CRISP-DM sehingga tahapan mulai dari pemahaman data hingga evaluasi model dapat berjalan secara terstruktur (Amalsyah et al., 2025).

Saat ini, ketersediaan data opini publik di media sosial sangat melimpah, namun proses pengolahannya menghadapi tantangan signifikan terkait karakteristik bahasa yang dinamis. Cuitan di platform X sering kali ditulis dengan struktur yang tidak baku, menggunakan singkatan, akronim, serta campuran bahasa asing yang sulit diproses oleh metode konvensional (Sintiya et al., 2025). Selain itu, keterbatasan panjang karakter memaksa pengguna memadatkan kalimat, yang sering kali menghilangkan konteks utuh dari sebuah opini. Metode analisis sentimen konvensional seperti *Naïve Bayes* atau SVM sering kali gagal menangkap nuansa kontekstual ini karena memperlakukan kata sebagai fitur yang berdiri sendiri, sehingga akurasi prediksi cenderung rendah pada kalimat yang kompleks (Lestari, 2025).

Pemahaman dependensi jangka panjang dan konteks kalimat dua arah menjadi keterbatasan metode konvensional, oleh karena itu perlunya penerapan algoritma *Deep Learning*. Dalam diskursus RUU Perampasan Aset, opini masyarakat sering kali disampaikan dengan gaya bahasa sindiran yang memerlukan pemahaman konteks menyeluruh. Oleh karena itu, arsitektur *Bidirectional Long Short-Term Memory* (BiLSTM) dipilih karena kemampuannya memproses informasi dari dua arah (maju dan mundur) secara simultan, sehingga makna kata dapat dipahami berdasarkan kata sebelum dan sesudahnya (Siswoyo et al., 2025).

Algoritma BiLSTM memerlukan masukan berupa vektor numerik yang merepresentasikan makna semantik kata. Salah satu pendekatan *embedding* yang banyak digunakan adalah *Word2Vec*, yang mampu menangkap hubungan semantik antar kata berdasarkan konteks kemunculannya dalam korpus (Saputra et al., 2025) (Yafik & Azhari, 2026). *Word2Vec* memiliki dua arsitektur utama, yaitu *Continuous Bag-of-Words* (CBOW) dan *Skip-gram*. CBOW mempelajari representasi kata dengan memprediksi kata target berdasarkan konteks sekelilingnya, sedangkan *Skip-gram* memprediksi kata konteks berdasarkan kata target.

Selain itu, tahapan *preprocessing* seperti *stemming* dan *stopword removal* juga berpotensi memengaruhi kualitas model analisis sentimen. Beberapa penelitian terbaru pada teks berbahasa Indonesia menunjukkan bahwa penerapan *stemming* dan *stopword removal* tidak selalu meningkatkan akurasi, bahkan pada kasus tertentu justru menurunkan performa klasifikasi sehingga efeknya perlu dievaluasi secara empiris pada setiap korpus yang digunakan (Pradana & Hayaty, 2019) (Angelina et al., 2023). Hal ini menjadi relevan dalam konteks cuitan mengenai RUU Perampasan Aset yang bersifat singkat, padat, dan sering kali menggunakan variasi bahasa tidak baku.

Di sisi lain, pemilihan teknik *word embedding* dan arsitektur model *deep learning* terbukti berperan penting dalam meningkatkan ketepatan klasifikasi sentimen. Studi-studi sebelumnya menunjukkan bahwa kombinasi *Word2Vec* dan BiLSTM mampu memberikan performa yang baik pada berbagai tugas analisis sentimen berbahasa Indonesia (Rustad et al., 2025). Namun, sebagian besar penelitian tersebut hanya menggunakan salah satu arsitektur *Word2Vec* tanpa membandingkan secara eksplisit perbedaan kinerja CBOW dan *Skip-gram*, serta belum mengkaji secara sistematis pengaruh variasi *stemming* dan *stopword removal* pada korpus opini kebijakan publik seperti RUU Perampasan Aset.

Selain itu, sebagian besar penelitian terdahulu juga masih terbatas pada analisis sentimen tanpa mengintegrasikan pemodelan topik untuk menggali konteks pembahasan secara lebih mendalam. Padahal, dengan menerapkan metode *Latent Dirichlet Allocation* setelah proses klasifikasi sentimen, peneliti dapat mengidentifikasi topik-topik dominan yang muncul pada masing-masing kategori

sentimen, sehingga menghasilkan analisis yang tidak hanya bersifat kuantitatif tetapi juga interpretatif.

Sejalan dengan temuan tersebut, penelitian ini tidak hanya membangun model BiLSTM berbasis pendekatan *Word2Vec* untuk analisis sentimen terhadap RUU Perampasan Aset, tetapi juga secara sistematis menguji dua faktor utama, yaitu skema *preprocessing* (dengan dan tanpa *stemming* serta *stopword removal*) dan pemilihan arsitektur *Word2Vec* (CBOW dan *Skip-gram*), serta mengintegrasikan pemodelan topik menggunakan LDA untuk menganalisis distribusi topik pada masing-masing sentimen.

Berdasarkan paparan tersebut, urgensi penelitian ini terletak pada pengembangan model analisis sentimen yang tidak hanya akurat tetapi juga adaptif terhadap karakteristik data media sosial yang informal dan sarat variasi bahasa. Sebagian besar penelitian terdahulu masih berfokus pada satu aspek saja, seperti pemilihan model atau teknik embedding, dan belum banyak yang secara komprehensif mengombinasikan pengujian skema *preprocessing* dan variasi arsitektur *Word2Vec* pada konteks kebijakan publik di Indonesia. Oleh karena itu, penelitian ini berupaya mengisi celah tersebut dengan membangun dan mengevaluasi model BiLSTM berbasis pendekatan *Word2Vec* yang diuji pada berbagai konfigurasi *preprocessing* dan arsitektur CBOW serta *Skip-gram*, serta dilengkapi dengan analisis topik menggunakan LDA, sehingga diharapkan mampu memberikan kontribusi empiris di bidang pemrosesan bahasa alami sekaligus menyediakan wawasan yang lebih terukur mengenai peta dukungan dan penolakan publik terhadap pengesahan RUU Perampasan Aset.

1.2 Rumusan Masalah

Berdasarkan latar belakang dan *gap* penelitian yang telah diuraikan, dirumuskan permasalahan, metode usulan, serta pertanyaan penelitian sebagai berikut.

1.2.1 Permasalahan

Media sosial X (Twitter) memegang peranan vital sebagai ruang publik digital tempat masyarakat menyuarakan aspirasi terkait Rancangan Undang-Undang (RUU) Perampasan Aset. Analisis sentimen terhadap opini-opini ini menawarkan nilai strategis untuk mengubah data kualitatif menjadi indikator tingkat penerimaan

publik yang terukur. Tanpa analisis yang sistematis, aspirasi masyarakat hanya akan menjadi tumpukan data yang sulit diterjemahkan menjadi wawasan kebijakan. Namun, dalam pengembangan model klasifikasi yang optimal, penelitian ini dihadapkan pada beberapa permasalahan teknis utama sebagai berikut:

1. Ketidakpastian skema *preprocessing* dalam penggunaan *stemming* dan *stopword removal* berpotensi menghilangkan informasi penting dalam teks, sehingga perlu diuji pengaruhnya terhadap kinerja model klasifikasi sentimen BiLSTM.
2. Ketidakjelasan pemilihan arsitektur *Word2Vec* apakah CBOW atau *Skip-gram* yang lebih optimal sebagai metode ekstraksi fitur pada model BiLSTM dalam merepresentasikan teks opini media sosial yang tidak baku.
3. Analisis sentimen hanya menunjukkan polaritas tanpa menjelaskan isu yang dibahas, sehingga diperlukan metode seperti *Latent Dirichlet Allocation* (LDA) untuk mengidentifikasi topik dominan dalam setiap sentimen.

1.2.2 Metode Usulan

Berdasarkan permasalahan yang telah dipaparkan, penelitian ini mengusulkan kerangka kerja berbasis CRISP-DM yang diintegrasikan dengan model *Bidirectional Long Short-Term Memory*. Proses diawali dengan crawling data dari media sosial X menggunakan kata kunci terkait “RUU Perampasan Aset”, kemudian dilakukan *preprocessing* teks yang mencakup *cleaning*, *case folding*, normalisasi kata tidak baku, *tokenizing*, *stopword removal*, dan *stemming*. Penelitian ini juga menguji beberapa variasi *preprocessing*, yaitu penggunaan *stopword removal* dan *stemming* secara bersamaan, penggunaan tanpa *stemming*, penggunaan *stemming* tanpa *stopword removal*, serta tanpa penggunaan keduanya.

Tahap pelabelan sentimen dilakukan menggunakan pendekatan *lexicon-based* dengan memanfaatkan *SentiWordNet*. Setiap token hasil *preprocessing* dicocokkan dengan kamus *SentiWordNet* untuk memperoleh skor polaritas positif dan negatif. Skor tersebut kemudian dijumlahkan untuk menentukan label sentimen pada setiap cuitan.

Sebagai pengembangan analisis, penelitian ini juga menerapkan metode *Latent Dirichlet Allocation* setelah tahap klasifikasi sentimen. LDA digunakan untuk mengekstraksi topik-topik dominan pada masing-masing kategori sentimen, sehingga tidak hanya diperoleh informasi polaritas opini, tetapi juga konteks isu yang mendasari dukungan maupun penolakan publik terhadap RUU Perampasan Aset.

Tahap akhir dilakukan evaluasi kinerja model menggunakan *confusion matrix* yang mencakup *accuracy*, *precision*, *recall*, dan *F1-Score* untuk mengukur performa klasifikasi sentimen. Selain itu, kualitas pemodelan topik menggunakan *Latent Dirichlet Allocation* dievaluasi menggunakan *coherence score* untuk menentukan jumlah topik yang paling optimal dan memastikan interpretasi topik yang dihasilkan memiliki tingkat konsistensi semantik yang baik.

1.2.3 Pertanyaan Penelitian

Dengan mempertimbangkan kasus yang telah diidentifikasi dan pendekatan yang telah diajukan, berikut adalah beberapa pertanyaan yang muncul:

1. Bagaimana pengaruh *stemming* dan *stopword removal* terhadap kinerja model BiLSTM berbasis *Word2Vec* dalam klasifikasi sentimen RUU Perampasan Aset pada platform X?
2. Bagaimana perbandingan kinerja arsitektur *Word2Vec* CBOW dan *Skip-gram* ketika digunakan sebagai metode ekstraksi fitur pada model BiLSTM untuk analisis sentimen RUU Perampasan Aset?
3. Bagaimana hasil pemodelan topik menggunakan *Latent Dirichlet Allocation* dalam mengidentifikasi topik dominan pada masing-masing kategori sentimen terhadap RUU Perampasan Aset?

1.3 Tujuan dan Manfaat Penelitian

Mengacu pada masalah dan pertanyaan yang telah diuraikan serta mempertimbangkan pendekatan yang telah diajukan, berikut merupakan tujuan dan manfaat dalam penelitian ini.

1.3.1 Tujuan Penelitian

Merujuk pada rumusan masalah dan pertanyaan penelitian yang telah dipaparkan, penelitian ini bertujuan untuk:

1. Mengetahui pengaruh penerapan tahapan *stemming* dan *stopword removal* terhadap kinerja model BiLSTM berbasis *Word2Vec* dalam klasifikasi sentimen RUU Perampasan Aset pada platform X.
2. Mengetahui perbandingan kinerja arsitektur *Word2Vec* (CBOW dan *Skip-gram*) ketika digunakan sebagai metode ekstraksi fitur pada model klasifikasi sentimen BiLSTM.
3. Mengidentifikasi topik dominan pada masing-masing kategori sentimen (positif dan negatif) terhadap RUU Perampasan Aset menggunakan metode pemodelan topik *Latent Dirichlet Allocation* (LDA).

1.3.2 Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat memberikan kontribusi, baik secara teoretis maupun praktis, sebagai berikut:

1. Memberikan bukti empiris mengenai pengaruh tahapan *stemming* dan *stopword removal* terhadap kinerja model BiLSTM berbasis *Word2Vec*.
2. Menyajikan analisis mengenai kinerja arsitektur *Word2Vec* CBOW dan *Skip-gram* dalam menangani data media sosial yang memiliki tingkat *noise* tinggi dan variasi bahasa tidak baku.
3. Memberikan kata interpretatif mengenai topik-topik dominan yang muncul pada masing-masing kategori sentimen melalui penerapan *Latent Dirichlet Allocation*.

1.4 Batasan Masalah

Agar penelitian ini terarah dan fokus, maka ruang lingkup penelitian dibatasi pada hal-hal berikut:

1. Objek penelitian dibatasi pada respons publik terhadap RUU Perampasan Aset yang dibahas pada platform media sosial X sebagai data utama.
2. Periode pengambilan data ditetapkan dalam rentang waktu Januari 2025 s.d. Oktober 2025.
3. Data yang dipakai berupa teks atau *tweet* berbahasa Indonesia. Penggunaan bahasa asing, campuran, atau simbol/emoji yang tidak relevan dalam *tweet* tidak menjadi fokus utama.
4. Penelitian berfokus pada penerapan algoritma *deep learning* menggunakan arsitektur BiLSTM berbasis pendekatan *Word2Vec*. Komparasi dilakukan

secara spesifik terhadap pengaruh penggunaan dan peniadaan *stopword removal* serta *stemming* pada tahap *preprocessing*, serta perbedaan kinerja arsitektur *Word2Vec* CBOW dan *Skip-gram* sebagai metode ekstraksi fitur.

5. Label klasifikasi sentimen dibatasi menjadi dua kelas kategori, yaitu sentimen positif dan sentimen negatif.
6. Hanya menggunakan 1 kata kunci pada saat pengambilan data yaitu “RUU Perampasan Aset”.

1.5 Sistematika Penulisan

Sistematika penulisan dari penelitian ini dapat dijelaskan sebagai berikut:

BAB I PENDAHULUAN

Bab ini memberikan penjelasan menyeluruh terkait latar belakang masalah mengenai urgensi analisis sentimen RUU Perampasan Aset, identifikasi dan perumusan masalah, tujuan dan manfaat penelitian, batasan masalah untuk menjaga fokus penelitian, serta sistematika penulisan yang digunakan.

BAB II TINJAUAN PUSTAKA

Bab ini mencakup kajian literatur terkait isu kebijakan RUU Perampasan Aset, konsep dasar analisis sentimen pada media sosial, kerangka kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM), tahapan *preprocessing*, metode pelabelan berbasis leksikon menggunakan *SentiWordNet*, metode ekstraksi fitur *Word2Vec* dengan arsitektur CBOW dan *Skip-gram*, arsitektur BiLSTM, metode pemodelan topik menggunakan *Latent Dirichlet Allocation*, teknik penanganan ketidakseimbangan data menggunakan SMOTE, serta metode evaluasi berbasis *confusion matrix*.

BAB III METODOLOGI PENELITIAN

Bab ini menguraikan tahapan teknis penelitian yang mengacu pada alur CRISP-DM, dimulai dari pemahaman bisnis terkait pemetaan isu RUU Perampasan Aset, pengumpulan dan pemahaman data cuitan dari platform X, *preprocessing* data, pelabelan sentimen menggunakan *SentiWordNet*, penerapan skenario *preprocessing*,

pembentukan representasi fitur menggunakan *Word2Vec* CBOW dan *Skip-gram*, penyeimbangan data latih menggunakan SMOTE, pembangunan model klasifikasi BiLSTM, hingga prosedur evaluasi model. Selain itu, bab ini juga menjelaskan penerapan *Latent Dirichlet Allocation* untuk mengidentifikasi topik dominan pada masing-masing kategori sentimen.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi penelitian yang meliputi proses pengolahan dataset, *preprocessing* data, pelatihan Model BiLSTM, serta evaluasi kinerja model. Selain itu, pada bab ini juga dilakukan analisis terhadap hasil klasifikasi sentimen yang diperoleh.

BAB V PENUTUP

Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan serta saran yang dapat digunakan sebagai bahan pengembangan penelitian selanjutnya.

PUSTAKA Daftar referensi yang digunakan dalam penelitian disertakan dalam bagian ini.