

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Pada era modern saat ini, *platform* jejaring sosial memegang peranan krusial dalam keseharian masyarakat, yang secara signifikan mengubah pola komunikasi, interaksi sosial, serta kebebasan berekspresi (Krisdiyanto, 2021). Kehadiran *platform* X (dahulu Twitter) berfungsi sebagai wadah publik berbasis digital yang memfasilitasi jutaan warganet untuk saling bertukar pandangan mengenai beragam fenomena sosial, dinamika politik, hingga merespons kebijakan pemerintah secara instan. X memiliki lebih dari 20 juta pengguna aktif di Indonesia, menjadikannya salah satu *platform* media sosial terbesar yang mencerminkan dinamika opini publik nasional (Naufal & Kusuma, 2022). *Platform* ini telah menjadi sumber data yang penting untuk analisis sentimen karena kebebasan ekspresi dan volume interaksi yang tinggi (Hasibuan dkk., 2024).

Kebebasan berpendapat yang difasilitasi oleh *platform* media sosial X menciptakan ruang demokratis bagi masyarakat untuk mengekspresikan pandangan mereka terhadap dinamika pemerintahan. Salah satu topik yang kini memicu gejolak opini publik secara masif adalah wacana kebijakan efisiensi anggaran sektor pendidikan pada tahun 2025. Isu ini memiliki tingkat urgensi dan sensitivitas yang tinggi karena menyangkut hajat hidup masyarakat luas serta keberlanjutan kualitas sumber daya manusia (Widangsa & Pratama, 2021). Meskipun Undang-Undang Dasar 1945 telah mengamanatkan alokasi anggaran pendidikan minimal 20% dari APBN, rencana efisiensi tersebut memunculkan spekulasi dan ketakutan publik terkait potensi pemotongan program-program strategis seperti dana Bantuan Operasional Sekolah (BOS), Program Indonesia Pintar (PIP), subsidi KIP Kuliah, hingga kesejahteraan tenaga pendidik (Prihantoro, 2024). Jika sentimen dan kepanikan masyarakat ini dibiarkan tanpa dianalisis, pemerintah akan kehilangan *insight* krusial mengenai titik kekhawatiran utama di akar rumput. Ketidakmampuan memetakan opini publik ini dapat berujung pada distorsi dan miskomunikasi kebijakan, memudarnya

tingkat kepercayaan masyarakat (*public trust*), hingga potensi penolakan sosial yang lebih luas saat kebijakan diimplementasikan (Dwiani dkk., 2025). Oleh karena itu, ekstraksi dan analisis sentimen secara *real-time* di ruang digital menjadi sangat esensial untuk dilakukan sebagai instrumen evaluasi dan *social listening* bagi para pemangku kebijakan (Romadhona & Isnain, 2024).

Platform X (dahulu Twitter) dipilih sebagai sumber data utama karena memiliki lebih dari 20 juta pengguna aktif di Indonesia dengan volume interaksi tinggi, fitur *hashtag* dan *trending topics* yang memfasilitasi diskusi publik *real-time*, serta aksesibilitas *crawling* yang mudah untuk data teks opini kebijakan (Naufal & Kusuma, 2022). Opini-opini yang terekspresikan di media sosial X memiliki potensi besar untuk dimanfaatkan sebagai sumber informasi dalam memahami persepsi dan sentimen masyarakat terhadap kebijakan publik. Analisis terhadap opini publik ini dapat memberikan wawasan berharga bagi pemerintah dan pembuat kebijakan untuk memahami respons masyarakat, mengidentifikasi area-area yang memerlukan perbaikan, dan merumuskan strategi komunikasi yang lebih efektif (Naraswati dkk., 2021). Pemanfaatan data media sosial untuk analisis kebijakan publik telah menjadi tren global yang semakin berkembang, khususnya dalam era *digital governance* dan *participatory democracy*, di mana pemerintah mulai memanfaatkan teknologi untuk memahami aspirasi rakyat (Adamansyah & Yudhistira, 2025).

Penggalian opini (*opinion mining*) atau yang sering disebut analisis sentimen, didefinisikan sebagai cabang keilmuan yang memadukan linguistik komputasi, kecerdasan buatan, serta pemrosesan bahasa alami. Penerapan disiplin ilmu ini bertujuan untuk membedah, menginterpretasi, serta mengelompokkan berbagai bentuk penilaian subjektif, sikap, maupun emosi yang tertuang di dalam sebuah data teks (Ma'rufudin & Yudhistira, 2025). Dalam konteks media sosial, analisis sentimen memungkinkan ekstraksi dan klasifikasi opini publik yang diekspresikan dalam bentuk *tweet*, komentar, atau *posting* menjadi kategori sentimen positif, negatif, atau netral. Teknik ini telah banyak diaplikasikan dalam berbagai domain, mulai dari analisis produk komersial, *monitoring brand reputation*, hingga evaluasi kebijakan publik dan kampanye politik, serta menjadi

semakin penting karena mampu mengolah data teks dalam jumlah besar secara otomatis dan memberikan *insight* yang dapat digunakan untuk pengambilan keputusan strategis (Fitrianti & Yudhistira, 2025).

Arsitektur jaringan *Long Short-Term Memory* (LSTM) secara khusus dirancang untuk mengatasi kendala *vanishing gradient* yang membatasi performa *Recurrent Neural Network* (RNN) konvensional dalam memproses sekuens data yang panjang. LSTM mampu mempertahankan konteks informasi dalam sekuens waktu yang lama melalui sistem gerbang (*gate mechanism*) yang terdiri dari *forget gate*, *input gate*, dan *output gate* untuk mengontrol aliran informasi di dalam sel memori (Farsiah dkk., 2022). LSTM memiliki kemampuan memodelkan dependensi jangka panjang antar kata dalam sebuah kalimat, sehingga ideal untuk tugas analisis sentimen teks media sosial yang sering kali mengandung konteks semantik yang kompleks dan urutan kata yang bervariasi (Farsiah dkk., 2022).

Keunggulan LSTM terlihat apabila dibandingkan dengan arsitektur *deep learning* lainnya seperti RNN, CNN, dan MLP. Berdasarkan penelitian Utami (2022), penggunaan RNN konvensional pada analisis sentimen teks hanya mampu mencapai akurasi sebesar 80% karena RNN murni sangat rentan mengalami hilangnya gradien (*vanishing gradient*) pada sekuens kalimat yang panjang. Di sisi lain, *Convolutional Neural Network* (CNN) terbukti cukup efektif untuk menangkap pola lokal (*n-gram*) dalam teks dengan akurasi mencapai 84% pada korpus berskala besar (Nurfauziyah dkk., 2025), namun arsitektur ini kurang optimal untuk memodelkan urutan panjang dan konteks global jika dibandingkan dengan model berbasis LSTM yang memproses teks secara sekuensial waktu-ke-waktu. Sementara itu, algoritma *Multi-Layer Perceptron* (MLP) terbukti memiliki keterbatasan karena tidak dirancang untuk data berurutan dan tidak memiliki memori internal atas posisi kata. Hal ini dibuktikan oleh Setiawan (2025) di mana klasifikasi sentimen menggunakan MLP hanya mampu mencapai akurasi 73,25%. Oleh karena itu, pada tugas analisis sentimen berbasis teks sekuensial, kombinasi *layer Embedding* dan memori LSTM cenderung lebih tangguh dan akurat dibandingkan MLP *feed-forward*.

Sebagai varian yang lebih sederhana dari LSTM, *Gated Recurrent Unit* (GRU) memiliki jumlah parameter yang lebih sedikit sehingga proses pelatihannya lebih cepat. Penelitian oleh Haerani (2026) memperlihatkan bahwa GRU mampu mencapai performa validasi yang sangat tinggi (94,61%) ketika dihadapkan pada *dataset* teks berskala masif. Meskipun demikian, berbagai studi menegaskan bahwa pada teks dengan konteks argumen yang rumit, dan multidimensional, struktur tiga gerbang pada LSTM sering kali menghasilkan pemahaman semantik yang lebih presisi dibandingkan GRU. Dalam konteks kebijakan pendidikan, penelitian Maulana dkk. (2023) membuktikan bahwa LSTM efektif digunakan untuk mengekstraksi opini publik terkait kebijakan Kurikulum Merdeka di media sosial. Oleh karena itu, penelitian ini memilih arsitektur LSTM satu arah yang dipadukan dengan *layer Embedding* karena pendekatan tersebut terbukti paling tangguh dalam menangkap konteks sekuensial *tweet* terkait efisiensi anggaran pendidikan yang sering memuat opini ambigu dan istilah kebijakan spesifik.

Penelitian ini menerapkan metode LSTM untuk analisis sentimen pada *platform* media sosial X terhadap isu efisiensi anggaran sektor pendidikan secara *end-to-end*. Tahapan penelitian meliputi pengumpulan data *tweet* melalui metode *crawling*, *preprocessing* teks (pembersihan, normalisasi, tokenisasi, dan *padding*), pembentukan representasi makna kata menggunakan *layer Embedding*, pelatihan model LSTM untuk mengklasifikasikan tiga kelas sentimen (Negatif, Netral, dan Positif), serta evaluasi kinerja akhir menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*. Merujuk pada evaluasi empiris terhadap arsitektur *deep learning* lainnya seperti RNN standar (H. Utami, 2022), CNN (Nurfauziyah dkk., 2025), MLP (Setiawan, 2025), dan GRU (Haerani, 2026), pemilihan LSTM dalam penelitian ini dilandasi oleh kemampuannya dalam menangani sekuens teks media sosial yang *noise* dan bergantung pada konteks. Implementasi sistem komputasi ini diharapkan dapat memberikan *insight* yang akurat mengenai tren opini publik terkait kebijakan efisiensi pendidikan.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang yang telah dijelaskan, maka perumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Bagaimana kinerja metode LSTM dalam mengklasifikasikan sentimen publik pada media sosial X terkait isu efisiensi anggaran sektor pendidikan, ditinjau dari nilai *accuracy*, *precision*, *recall*, dan *f1-score*?
2. Bagaimana kecenderungan sentimen masyarakat (dominan positif, negatif, atau netral) serta topik bahasan utama yang muncul dalam opini publik mengenai kebijakan efisiensi anggaran pendidikan tersebut?

1.3 Tujuan

Penelitian ini bertujuan untuk menjawab rumusan masalah yang telah dirumuskan, sehingga pelaksanaannya dapat sesuai dengan tujuan yang telah ditetapkan. Adapun tujuan yang ingin dicapai dalam penelitian ini meliputi.

1. Mengimplementasikan metode LSTM untuk menghasilkan klasifikasi sentimen publik (positif, negatif, netral) di media sosial X terhadap isu efisiensi anggaran sektor pendidikan, serta mengevaluasi performa model tersebut menggunakan metrik *accuracy*, *precision*, *recall*, dan *f1-score*.
2. Menganalisis kecenderungan sentimen masyarakat (dominan positif, negatif, atau netral) serta mengidentifikasi topik bahasan utama (*topic modelling*) yang muncul dalam opini publik mengenai kebijakan efisiensi anggaran pendidikan tersebut untuk memberikan gambaran persepsi masyarakat.

1.4 Batasan Masalah

Penelitian ini akan fokus pada.

1. Sumber Data: Data teks berbahasa Indonesia yang diperoleh dari media sosial X menggunakan teknik *crawling*, dengan kata kunci pencarian: “efisiensi anggaran pendidikan”, “pemotongan anggaran pendidikan”, “dana pendidikan dipotong”, serta kombinasi terkait “dana BOS”, “PIP”, atau “KIP Kuliah”.
2. Periode Pengambilan Data: Rentang waktu *crawling data* dibatasi mulai tanggal 1 Februari 2025 hingga 31 Maret 2025. Pemilihan batasan waktu ini

didasarkan pada tingginya intensitas perbincangan (*peak momentum*) warganet di media sosial X terkait wacana kebijakan efisiensi anggaran pendidikan, sehingga data yang diperoleh mampu merepresentasikan gejolak opini publik yang paling otentik dan relevan.

3. Volume Data: Data teks yang digunakan dalam penelitian ini dibatasi sebanyak 1.500 tweet berbahasa Indonesia. Data tersebut dikumpulkan secara acak dan representatif, yang selanjutnya akan melalui tahapan *preprocessing* serta pelabelan kelas sentimen.
4. Metode Algoritma: Algoritma utama yang digunakan adalah *Long Short Term Memory* (LSTM) untuk klasifikasi sentimen tiga kelas (positif, negatif, netral), dengan penggunaan lapisan *Word Embedding* untuk ekstraksi fitur semantik teks.
5. Teknik *Preprocessing*: Proses pembersihan dan pengolahan teks dilakukan menggunakan bahasa pemrograman Python, mencakup: *Cleaning Data* (pembersihan simbol/url), *Case Folding*, *Tokenizing Data*, dan *Normalizing Data* (Menormalkan kata-kata *slang* atau singkatan).
6. Tools & Implementasi: Pembangunan model dilakukan menggunakan *library* TensorFlow/Keras. Visualisasi hasil dan antarmuka pengguna (*user interface*) dibangun menggunakan *framework* Streamlit agar sistem dapat diakses melalui *web dashboard*.
7. Ruang Lingkup Analisis: Fokus penelitian terbatas pada analisis teks (*text mining*) dan tidak mencakup analisis terhadap konten multimedia (gambar, video, atau audio) maupun analisis akun *bot/buzzer*.

1.5 Manfaat

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut.

1. Bagi Pemerintah: Memberikan *dashboard* analisis sentimen berbasis *deep learning* untuk membantu perumusan kebijakan pendidikan yang lebih responsif terhadap aspirasi masyarakat dengan pemahaman konteks yang lebih mendalam.

2. Bagi Akademisi: Menjadi referensi penelitian terkait implementasi metode *Long Short Term Memory* (LSTM) dalam analisis sentimen kebijakan publik, khususnya di sektor pendidikan, serta kontribusi dalam pengembangan teknik *deep learning* untuk *natural language processing* berbahasa Indonesia.
3. Bagi Masyarakat: Meningkatkan transparansi pengelolaan anggaran pendidikan melalui analisis opini publik yang objektif, sistematis, dan berbasis teknologi *artificial intelligence* yang dapat memberikan *insights* yang lebih akurat.