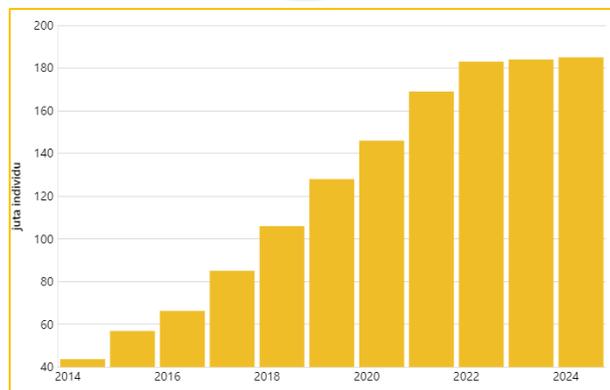


# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Kemudahan masyarakat dalam mengakses internet masa sekarang menunjukkan perkembangan teknologi informasi dan komunikasi, dengan kemudahan ini publik dapat mengakses media sosial untuk berkomunikasi dan memperoleh berbagai informasi. Media sosial sebagai sarana komunikasi berfungsi untuk menyebarkan informasi secara cepat dan luas, selain itu sebagai sarana perdagangan, penyebarluasan kebijakan pemerintah, kampanye politik, dan dakwah agama [1]. Oleh karena itu, kerap menemukan para pengguna media sosial yang berkomentar atau membuat postingan mengenai informasi atas suatu peristiwa. Kebebasan dalam berkomentar ini membuat setiap warga negara berhak mengeluarkan pendapatnya secara bebas dan bertanggung jawab.

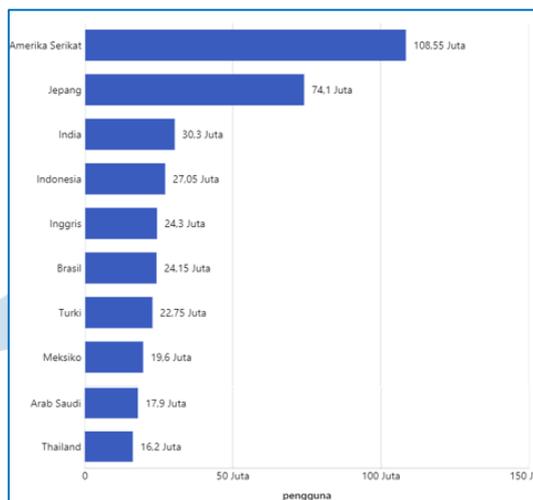


Gambar 1.1 Survei Pengguna Sosial Media di Indonesia 2024 oleh *We Are Social*

Hasil survei *We Are Social* pada Gambar 1.1 menunjukkan bahwa adanya peningkatan penggunaan media sosial di Indonesia dari tahun ke tahun. Sesuai pernyataan *We Are Social*, di awal tahun 2024 terdapat 185 juta jiwa pemakai internet di Indonesia, sebesar 66.5% dari keseluruhan penduduk nasional berjumlah 278.7 juta jiwa. Pemakai fasilitas internet terekam bertambah sekitar 1.5 juta orang

yang berarti meningkat 0.8% di Indonesia awal tahun 2024 dari Januari 2023 [2]. Hal ini membuktikan bahwa, meningkatnya popularitas penggunaan media sosial seperti X, Facebook, dan Instagram seiring dengan berkembangnya pengguna internet yang dapat mengakses informasi dari seluruh dunia. Informasi yang tersebar di media sosial bisa menjadi viral jika menarik perhatian masyarakat umum [3].

Pada Gambar 1.2 menunjukkan terpadat sekitar 27,5 juta pengguna X di Indonesia per Oktober 2023, jumlah itu menempatkan Indonesia di peringkat keempat global [4]. Penelitian ini akan menggunakan X sebagai sarana penggalian informasi karena dianggap salah satu media sosial terpopuler dan yang erat kaitannya dengan isu-isu viral yang sedang berlangsung dan dapat dimanfaatkan untuk memperoleh informasi suatu topik permasalahan, tak terkecuali tentang kinerja pemerintahan [5]. Pemerintahan di momen ini juga tak lolos dari komentar masyarakat di media sosial X. Pemerintahan sekarang yang dipimpin Presiden Joko Widodo memperoleh bermacam komentar, mulai dari kekaguman, kecaman, usulan, satir hingga ujaran kemarahan.



Gambar 1.2 Jumlah Pengguna X di dunia 2023 oleh laporan *We Are Social*

Pembangunan infrastruktur menjadi program prioritas di masa kepemimpinannya yang pertama dan periode kedua fokus pada pembangunan dan peningkatan kapasitas SDM [6]. Pemerintahan Jokowi pada periode kedua memiliki lima prioritas kerja, yaitu "Pembangunan SDM", "Pembangunan Infrastruktur", "Pemangkasan Regulasi", "Reformasi Birokrasi", dan "Transformasi Ekonomi". Proses klasifikasi akan dilakukan berdasarkan *tweet* publik dengan *keyword* lima prioritas kerja tersebut dan *keyword* lainnya, seperti "Presiden Jokowi", "Kerja Jokowi", "Prestasi Jokowi", "Proyek Jokowi", "Pemerintahan Jokowi", "Menteri Jokowi", "Kinerja Jokowi", "Kinerja Pemerintah", "Kabinet Jokowi", "Kabinet Indonesia Maju", "Menuju Indonesia Maju", "NKRI Membangun", "Kerja Menteri", "Ibu Kota Nusantara", "Hukum Indonesia", "Politik Indonesia", "Pendidikan Indonesia", "Ekonomi Indonesia", "Komisi Pemberantasan Korupsi", "Kesejahteraan Rakyat", "Kartu Indonesia Sehat", dan "Kartu Indonesia Pintar".

Salah satu metode yang dapat digunakan untuk menganalisis ragam opini *tweet* adalah analisis sentimen. Analisis sentimen dilakukan dengan bantuan algoritma *Natural Language Processing* untuk memproses dan menganalisis *tweet* secara otomatis. Metode klasifikasi yang digunakan dalam penelitian ini adalah *Long Short-Term Memory* (LSTM) dan *Convolutional Neural Network* (CNN). Pelabelan data menggunakan kamus *InSet Lexicon* untuk mengelompokkan *tweet* ke dalam kelas positif dan negatif serta perhitungan skor sentimen setiap *tweet*. Penelitian ini menggunakan bahasa pemrograman Python dengan tools Visual Studio Code. Penelitian ini akan membandingkan dua metode pembobotan kata, yaitu *Term Frequency Inverse Document* (TF-IDF) dan *Word Embedding* untuk meningkatkan kemampuan model memahami dan mengolah bahasa alami. Selain itu, penelitian ini juga akan membandingkan tujuh *optimizer*, diantaranya *Adam*, *RMSprop*, *SGD*, *Adagrad*, *Adadelta*, *Adamax*, dan *Nadam*. *Optimizer* bertujuan untuk meminimalkan *loss function* agar model menjadi lebih akurat. *Cross Industry*

*Standard Process of Data Mining* (CRISP-DM) digunakan sebagai kerangka kerja atau pedoman dalam analisis data dengan metode LSTM dan CNN.

Penelitian yang telah dilakukan oleh Fitroh dan F. Huda untuk mengetahui tren penelitian analisis sentimen berbasis *deep learning*, ditemukan bahwa metode terpopuler di tahun 2020-2023 adalah LSTM dan CNN dengan akurasi tertinggi mencapai 99% dan rata-rata 89% [7]. Pada penelitian oleh H. Tsaniya, R. Rosadi and A. S. Abdullah dengan topik serupa dengan penelitian ini, yaitu pemerintahan Jokowi menggunakan metode CNN didapatkan akurasi 57% [8]. Penelitian lain oleh A. F. Hidayatullah, S. Cahya, dengan topik pemilu dengan didapatkan akurasi LSTM 84.20%, akurasi CNN 84.05%, akurasi CNN+LSTM 84.30%, akurasi GRU+LSTM 84.50%, dan akurasi Bidirectional LSTM 84.60% [9]. Penelitian lain oleh F. M. Cholis, M. C. Utomo. dan N. R. Fadhlina terkait analisis sentimen data X isu pemerintah khususnya pemilu didapatkan akurasi terbaik oleh metode LSTM mencapai 92% [10].

Perbedaan penelitian ini dengan penelitian sebelumnya adalah pada objek penelitian termasuk *keyword* yang dipilih, periode data penelitian, dan metode penelitian. Objek penelitian ini adalah sentimen masyarakat Indonesia terhadap pemerintahan Jokowi dari Oktober 2014 hingga April 2024. Pengambilan data dengan cara *web scraping* pada X menggunakan *Tweet-harvest* untuk mendapatkan data melalui *keyword*. Pada tahap *normalization* penelitian ini, menggunakan tiga file berisikan *list* kata bahasa baku dan tidak baku Indonesia, terdiri atas "new\_kamusalay.csv", "colloquial-indonesian-lexicon.csv", dan "kbba.txt", dilanjutkan *sentiment labeling* menggunakan *InSet Lexicon*, seluruh file bersumber dari github. Melakukan analisis dengan metode LSTM dan CNN menggunakan dua fitur ekstraksi, yaitu TF-IDF dan *Word Embedding* serta membandingkan tujuh *optimizer*, yaitu *Adam*, *RMSprop*, *SGD*, *Adagrad*, *Adadelta*, *Adamax*, dan *Nadam*. Pada tahap terakhir penelitian ini menggunakan *confusion matrix* untuk evaluasi

kinerja model LSTM dan CNN yang telah dilakukan untuk menghasilkan nilai akurasi, presisi, *recall*, dan *f1-score*.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang sudah diuraikan diatas maka dapat dibuat rumusan masalah sebagai berikut:

1. Bagaimana gambaran lebih jelas mengenai respon publik terhadap prioritas kerja pemerintahan Jokowi pada sosial media X?
2. Bagaimana performa metode LSTM pada data tentang pemerintahan Jokowi menggunakan fitur ekstraksi TF-IDF dan *Word Embedding* dengan *optimizer Adam, RMSprop, SGD, Adagrad, Adadelata, Adamax, dan Nadam*?
3. Bagaimana performa metode CNN pada data tentang pemerintahan Jokowi menggunakan fitur ekstraksi TF-IDF dan *Word Embedding* dengan *optimizer Adam, RMSprop, SGD, Adagrad, Adadelata, Adamax, dan Nadam*?

## 1.3 Batasan Masalah

Beberapa batasan masalah ditetapkan agar pembahasan pada penelitian ini dapat tetap terarah. Batasan masalah tersebut antara lain:

1. *Tweet* yang digunakan untuk analisis sentimen adalah *tweet* yang berbahasa Indonesia.
2. *Tweet* yang berhubungan dengan pemerintahan Jokowi diantaranya prioritas kerja Jokowi dan kinerja pemerintahan Jokowi.
3. Pelabelan data menggunakan *InSet Lexicon* yang terdiri atas label positif dan negatif.
4. Penelitian menggunakan metode LSTM dan CNN.

## 1.4 Tujuan dan Manfaat Penelitian

### 1.4.1 Tujuan Penelitian

Adapun tujuan dari penelitian ini, sebagai berikut:

1. Mengidentifikasi pola sentimen atas respon publik yang lebih jelas terhadap prioritas kerja pemerintahan Jokowi pada sosial media X.
2. Mengevaluasi dan membandingkan performa metode LSTM dalam analisis data *tweet* tentang pemerintahan Jokowi menggunakan fitur ekstraksi TF-IDF dan *Word Embedding* dengan *optimizer Adam, RMSprop, SGD, Adagrad, Adadelta, Adamax, dan Nadam.*
3. Mengevaluasi dan membandingkan performa metode CNN dalam analisis data *tweet* tentang pemerintahan Jokowi menggunakan fitur ekstraksi TF-IDF dan *Word Embedding* dengan *optimizer Adam, RMSprop, SGD, Adagrad, Adadelta, Adamax, dan Nadam.*

### 1.4.2 Manfaat Penelitian

Adapun manfaat dari penelitian ini, sebagai berikut:

1. Memberikan pemahaman yang lebih jelas tentang respon publik terhadap prioritas kerja pemerintahan Jokowi pada sosial media X.
2. Meningkatkan pemahaman tentang kinerja dan keefektifan metode LSTM dalam menganalisis *tweet* tentang pemerintahan Jokowi menggunakan fitur ekstraksi TF-IDF dan *Word Embedding* dengan *optimizer Adam, RMSprop, SGD, Adagrad, Adadelta, Adamax, dan Nadam.*

3. Meningkatkan pemahaman tentang kinerja dan keefektifan metode CNN dalam menganalisis *tweet* tentang pemerintahan Jokowi menggunakan fitur ekstraksi TF-IDF dan *Word Embedding* dengan *optimizer Adam, RMSprop, SGD, Adagrad, Adadelta, Adamax, dan Nadam*.

## 1.5 Sistematika Penulisan

Proses penulisan laporan penelitian ini disusun menjadi lima bagian berbeda, masing-masing diuraikan sebagai berikut:

### BAB 1 – PENDAHULUAN

Pada bagian ini memberikan penjelasan yang melatarbelakangi topik yang akan dibahas pada penelitian ini. Alasannya diuraikan dalam urutan sebagai berikut: perumusan masalah, keterbatasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

### BAB 2 – LANDASAN TEORI

Penelitian terdahulu, Analisis Sentimen, *Deep Learning, Natural Language Processing (NLP), Web Scrapping, Cross Industry Standard Process for Data Mining (CRISP-DM), Long Short-Term Memory, Convolutional Neural Network, Feature Extraction, TF-IDF, Word Embedding, Confusion Matrix, X, Python, Visual Studio Code, dan Tweet-harvest* adalah konsep teoritis yang berfungsi sebagai landasan teoritis untuk penelitian ini.

### BAB 3 – METODOLOGI PENELITIAN

Pada bagian ini mencakup penjelasan metode-metode yang akan digunakan dan digambarkan dalam bentuk diagram alur untuk menguraikan proses berurutan dari penelitian ini. Bagian ini terdiri dari: gambaran umum objek penelitian, metode penelitian, teknik pengumpulan data, teknik

analisis data menggunakan CRISP-DM (*Business Understanding, Data Understanding, Data Preparation, Modeling, dan Evaluation*).

#### BAB 4 – HASIL DAN PEMBAHASAN

Pada bagian ini mencakup temuan penelitian yang dilakukan pada implementasi model menggunakan *Long Short-Term Memory* dan *Convolutional Neural Network* dengan membandingkan fitur ekstraksi TF-IDF dan *Word Embedding* juga membandingkan *optimizer Adam, RMSprop, SGD, Adagrad, Adadelta, Adamax, dan Nadam*. Evaluasi kinerja model menggunakan *confusion matrix* kemudian dibuatkan *classification report* dari kedua model terbaik untuk menentukan hasil temuan terbaik.

#### BAB 5 – SIMPULAN DAN SARAN

Pada bagian ini, berisi simpulan yang memberikan jawaban atas rumusan masalah penelitian dan tujuan penelitian yang telah dibuat, beserta informasi tambahan yang didapatkan dari hasil penelitian. Bab ini juga berisi saran dari penulis untuk pengembangan penelitian dari segi teknik atau metode yang digunakan untuk hasil yang lebih baik di kemudian hari.

