

diselesaikan, menjelaskan relevansi dan pentingnya penelitian ini, serta memberikan gambaran umum mengenai struktur tesis atau skripsi.

- Bab 2 LANDASAN TEORI
Bab ini akan berisi tentang tinjauan teori yang akan menjadi bahan-bahan pendukung dalam pengerjaan skripsi.
- Bab 3 METODOLOGI PENELITIAN
Bab ini akan menjelaskan tentang metode yang akan diimplementasi dalam melakukan penelitian dalam skripsi
- Bab 4 HASIL DAN DISKUSI
Bab ini akan menunjukkan hasil-hasil yang diperoleh dari penelitian dan analisis yang telah dilakukan.
- Bab 5 KESIMPULAN DAN SARAN
Bab ini akan berisi tentang rangkuman tentang keseluruhan penelitian dan hasil yang didapatkan. Akan juga berisi tentang saran untuk penelitian kedepannya.



BAB 2

LANDASAN TEORI

2.1 Komisi Pemilihan Umum

KPU, singkatan dari Komisi Pemilihan Umum, adalah lembaga independen di Indonesia yang bertugas mengatur, menyelenggarakan, dan mengawasi pemilihan umum di tingkat nasional, provinsi, dan kabupaten/kota, sebagaimana diatur dalam Undang-Undang Nomor 15 Tahun 2011 tentang Penyelenggara Pemilihan Umum [13]. Penelitian yang dilakukan oleh [14] menyoroti transformasi KPU Indonesia dari institusi teknis menjadi institusi politik yang semakin berperan dalam proses demokratisasi. Penelitian tersebut menekankan perubahan signifikan dalam peran KPU seiring dengan perubahan konstitusi, peraturan perundang-undangan, serta dinamika politik dalam negeri. Selain itu, penelitian oleh [15] juga memberikan perspektif yang mendalam tentang tantangan dan upaya reformasi yang dihadapi oleh KPU dalam menjaga integritas, transparansi, dan akuntabilitas dalam penyelenggaraan pemilihan umum. Keseluruhan, telaah literatur ini memberikan pemahaman yang komprehensif tentang peran serta, tantangan, dan perubahan dalam kinerja KPU Indonesia dalam konteks proses demokratisasi di Indonesia.

2.2 Analisis Sentimen

Analisis sentimen merupakan salah satu aspek penting dalam text mining yang bertujuan untuk mengidentifikasi pendapat serta tingkat subjektivitas individu terhadap suatu topik, peristiwa, atau masalah tertentu [6]. Pusat perhatian dalam analisis sentimen adalah memproses pendapat yang dapat diklasifikasikan sebagai memiliki nilai sentimen positif atau negatif. Proses analisis sentimen melibatkan pengkajian terhadap berbagai macam pendapat, penilaian, evaluasi, sentimen, sikap, dan emosi yang berkaitan dengan entitas, produk, atau peristiwa yang sedang dibahas [16].

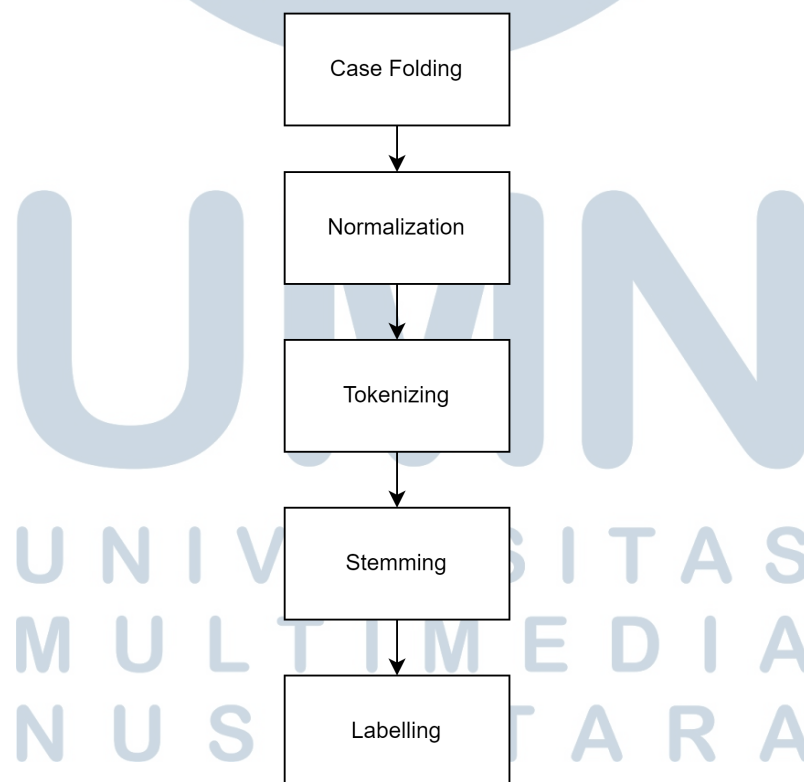
2.3 X

X, yang sebelumnya dikenal sebagai Twitter, adalah platform media sosial yang telah menjadi pusat perhatian dalam berbagai penelitian terkait analisis teks

dan pemahaman sosial. Dalam lima tahun terakhir, penelitian telah menyoroti beragam aspek dari platform ini, mulai dari pola perilaku pengguna hingga penyebaran informasi dan sentimen dalam konteks sosial dan politik. Misalnya, penelitian oleh Wang et al [17]. mengungkapkan pentingnya identifikasi dan analisis opini pengguna dalam topik-topik tertentu, sementara studi oleh Liu et al[18]. menyoroti peran penting Twitter dalam mendeteksi tren dan peristiwa penting secara real-time. Dengan demikian, X (formerly Twitter) terus menjadi subjek penelitian yang menarik dalam pemahaman perilaku dan dinamika sosial dalam era digital yang terus berkembang.

2.4 Text pre-Processing

Text preprocessing adalah tahapan penting dalam analisis sentimen yang bertujuan untuk membersihkan dan mengubah teks mentah menjadi bentuk yang lebih terstruktur dan siap untuk dianalisis. Berikut adalah beberapa langkah umum dalam text preprocessing [1].



Gambar 2.1. Text pre-Processing flowchart [1]

- *Case Folding*, merupakan tahapan dalam pengolahan teks yang mengubah semua huruf menjadi huruf kecil atau huruf besar agar dapat diolah secara seragam tanpa memperhatikan perbedaan ukuran huruf. Hal ini bertujuan untuk menyamakan representasi teks yang memiliki variasi dalam penggunaan huruf kapitalisasi sehingga memudahkan dalam analisis dan pengolahan teks secara konsisten [19].
- *Normalization*, merupakan tahap yang bertujuan untuk menyamakan atau mengubah teks menjadi format standar yang lebih mudah untuk diproses, seperti menghilangkan tanda baca, mengganti karakter khusus dengan karakter yang sesuai, atau mengkonversi angka menjadi representasi teks yang konsisten. Proses normalisasi membantu mengurangi variasi dalam teks dan mempersiapkan data untuk analisis atau pemrosesan lebih lanjut dengan memastikan konsistensi dalam representasi teks [20].
- *Tokenization* merupakan tahap menguraikan teks menjadi unit-unit kecil yang disebut sebagai "token". Token dapat berupa kata, frasa, atau karakter terpisah, tergantung pada jenis tokenisasi yang digunakan. Tujuan utama dari tokenisasi adalah untuk membagi teks menjadi bagian-bagian yang lebih kecil dan bermakna sehingga dapat diolah lebih lanjut dalam analisis atau pemrosesan teks. Misalnya, dalam tokenisasi kata, teks akan dibagi menjadi kata-kata individual, sedangkan dalam tokenisasi karakter, teks akan dibagi menjadi karakter-karakter terpisah [21].
- *Stemming* merupakan proses yang bertujuan untuk mengubah kata-kata menjadi bentuk dasar atau kata-kata akarnya. Tujuan dari stemming adalah untuk mengurangi variasi dalam kata-kata yang memiliki akar yang sama tetapi ditulis dengan cara yang berbeda. Misalnya, kata "berlari", "berlari-lari", dan "berlari-lah" akan diubah menjadi bentuk dasarnya "lari" [22].
- *Labelling* merupakan proses memberikan klasifikasi atau label pada teks atau dokumen berdasarkan ekspresi emosional yang terkandung di dalamnya, seperti positif, negatif, atau netral [23].

2.5 Artificial Neural Network

Artificial Neural Network(ANN) merupakan salah satu *framework machine learning* yang sering digunakan. ANN merupakan model yang terinspirasi dari

sistem syaraf pada makhluk hidup. Struktur ANN bekerja dengan cara yang mirip dengan sistem syaraf biologis pada hewan. Pertama-tama, input diberikan kepada jaringan saraf, yang kemudian diteruskan melalui berbagai lapisan neuron, termasuk lapisan input, tersembunyi, dan output. Di setiap neuron, input yang diterima dijumlahkan dengan bobot yang sesuai, dan hasilnya dilewati melalui fungsi aktivasi untuk menentukan apakah neuron tersebut akan "aktif" atau tidak. Proses ini berulang untuk setiap neuron dalam jaringan, dan output akhir diperoleh dari lapisan output. Sebelum dapat digunakan, jaringan harus dipelajari dari data yang ada melalui proses pelatihan di mana bobotnya disesuaikan untuk meminimalkan kesalahan prediksi. Setelah pelatihan selesai, jaringan dapat digunakan untuk membuat prediksi atau klasifikasi berdasarkan input baru yang diberikan padanya. Dengan demikian, ANN memungkinkan pembelajaran dari data, pemodelan pola yang kompleks, dan pengambilan keputusan berdasarkan pola-pola tersebut. [24]

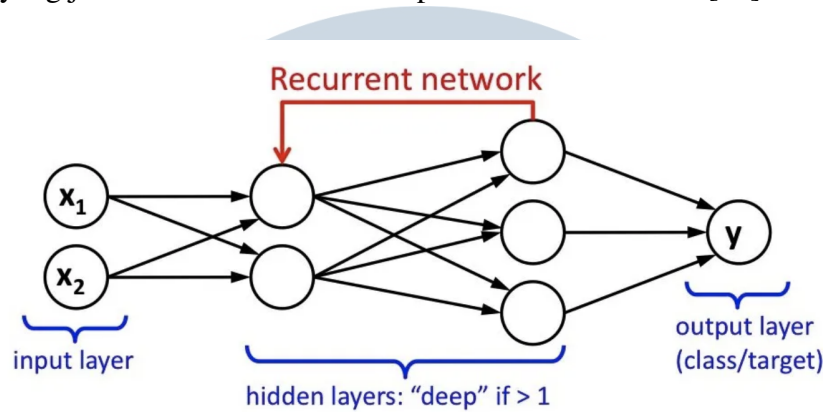
Salah Satu bentuk neural network terdasar adalah *Multilayer Perceptron* (MLP). MLP merupakan sebuah *feedforward artificial neural network* yang terdiri minimal 3 layer: *input layer*, 1 atau lebih *hidden layers*, dan 1 *output layer*. Setiap layer memiliki beberapa neuron yang memiliki ikatan antara neuron pada layer selanjutnya. Dasar dari rumus matematika MLP berhubungan dengan menghitung jumlah bobot dari input yang kemudian akan melewati fungsi aktivasi seperti sigmoid, tanh, or ReLU (Rectified Linear Unit)[25].

$$y = f \left(\sum_{i=1}^n w_i x_i + b \right) \quad [25] \quad (2.1)$$

dimana y adalah output, f adalah fungsi aktivasi, w_i adalah bobot dari input, x_i adalah inputs, dan b adalah bias.

Artificial Neural Network memiliki beberapa metode yang dikembangkan salah satunya *Recurrent Neural Network*, Jaringan neural berulang (RNN) adalah jenis model dalam bidang deep learning yang terlatih untuk memproses dan mengubah data input berurutan menjadi output berurutan yang spesifik. Data berurutan ini bisa berupa teks, urutan waktu, atau entitas berurutan lainnya di mana komponennya saling terhubung secara semantik dan mengikuti aturan sintaksis. RNN berfungsi sebagai sistem perangkat lunak kompleks yang terdiri dari beberapa komponen yang saling terhubung, meniru cara manusia melakukan transformasi data berurutan, seperti menerjemahkan teks dari satu bahasa ke bahasa lain. Namun, sebagian besar RNN saat ini telah tergantikan oleh pendekatan berbasis transformer

dan model bahasa besar (LLM) dalam kecerdasan buatan (AI), yang menawarkan efisiensi yang jauh lebih baik dalam memproses data berurutan. [26]



Gambar 2.2. Recurrent Neural Network [26]

Jaringan Neural Berulang (RNN) terbentuk dari unit pemrosesan data yang disebut neuron, yang bekerja secara kolaboratif untuk menangani tugas-tugas yang kompleks. Neuron-neuron ini diposisikan dalam tiga bagian: sebagai input, output, dan lapisan tersembunyi. Lapisan input menerima informasi yang akan diproses, sementara lapisan output menghasilkan output akhir. Proses analisis data dan prediksi berlangsung di dalam lapisan tersembunyi.

2.6 Algoritma Long Short Term Memory

Long Short-Term Memory (LSTM) merupakan sebuah jenis dari recurrent neural network (RNN) yang dirancang untuk mengatasi masalah vanishing gradient dan exploding gradient yang sering terjadi pada RNN konvensional [27]. LSTM memungkinkan jaringan untuk mengingat informasi dalam jangka waktu yang lama, sehingga sangat cocok digunakan pada tugas-tugas yang melibatkan urutan data seperti teks, audio, dan video.

Model LSTM terdiri dari beberapa komponen utama, yaitu forget gate, input gate, output gate, dan sel memori. Berikut adalah rumus-rumus yang menggambarkan aliran data melalui LSTM [28, 29]:

- **Forget Gate (f_t)**

Forget gate digunakan untuk mengatur seberapa banyak informasi dari sel memori sebelumnya akan dilupakan. Ini dilakukan dengan menggunakan fungsi sigmoid untuk menghasilkan nilai antara 0 dan 1 untuk setiap elemen

dalam sel memori, di mana nilai yang mendekati 1 menunjukkan bahwa informasi tersebut harus dipertahankan, sedangkan nilai yang mendekati 0 menunjukkan bahwa informasi tersebut harus dilupakan [28].

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2.2)$$

- **Input Gate (i_t)**

Input gate menentukan seberapa banyak informasi baru akan dimasukkan ke dalam sel memori. Pertama, fungsi sigmoid digunakan untuk menghasilkan vektor yang menentukan seberapa banyak setiap elemen dalam sel memori harus diperbarui. Selanjutnya, nilai kandidat baru \tilde{C}_t dihitung menggunakan fungsi tanh, yang menghasilkan nilai antara -1 dan 1, yang mewakili informasi yang diusulkan untuk dimasukkan ke dalam sel memori [29].

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2.3)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (2.4)$$

- **Update Cell State (C_t)**

Setelah menghitung input gate dan nilai kandidat baru, sel memori diperbarui dengan menggunakan kedua informasi tersebut. Forget gate digunakan untuk menentukan seberapa banyak informasi lama harus dipertahankan, sementara input gate digunakan untuk menentukan seberapa banyak informasi baru harus dimasukkan [30].

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (2.5)$$

- **Output Gate (o_t)**

Output gate mengontrol output yang dihasilkan oleh LSTM pada waktu t . Ini dilakukan dengan menggunakan fungsi sigmoid untuk menentukan seberapa banyak informasi dalam sel memori harus diekstraksi sebagai output. Selanjutnya, output dari LSTM pada waktu t dihitung sebagai hasil dari perkalian antara output gate dan fungsi tanh dari sel memori yang diperbarui [31].

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (2.6)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (2.7)$$

Dalam rumus-rumus di atas, x_t adalah input pada waktu t , h_{t-1} adalah output pada waktu sebelumnya, W dan b masing-masing adalah matriks bobot dan bias untuk setiap gate, dan σ adalah fungsi sigmoid.

Mengimplementasikan LSTM sebagai model melibatkan beberapa langkah kunci. Pertama, arsitektur jaringan harus didefinisikan, dengan menentukan jumlah lapisan LSTM dan jumlah unit per lapisan. Selanjutnya, jaringan diinisialisasi dengan bobot dan bias acak. Selama pelatihan, LSTM diberi urutan data input, dan outputnya dibandingkan dengan label sebenarnya untuk menghitung fungsi kerugian, seperti mean squared error atau cross-entropy loss. Algoritma backpropagation through time (BPTT) kemudian digunakan untuk menyesuaikan bobot dan bias dengan menghitung gradien dari fungsi kerugian terhadap setiap bobot dan bias, dan memperbaruinya ke arah yang mengurangi kerugian. Proses ini diulangi selama banyak iterasi, atau epoch, hingga jaringan mencapai kinerja yang memuaskan. Setelah dilatih, LSTM dapat digunakan untuk berbagai tugas seperti peramalan deret waktu, pemodelan bahasa, dan klasifikasi urutan, yang menunjukkan fleksibilitas dan kekuatan model pembelajaran mendalam dalam menangani data berurutan.

2.7 Confusion Matrix

Confusion Matrix adalah tabel yang mencerminkan hasil klasifikasi, menunjukkan jumlah data uji yang diklasifikasikan dengan benar dan yang salah. Metode ini merupakan cara yang efektif untuk mengevaluasi model yang telah dilatih, karena matriks kebingungan dapat digunakan pada masalah klasifikasi biner maupun klasifikasi multikelas [32].

Tabel 2.1. Confusion Matrix

		Prediksi	
		Positif	Negatif
Aktual	Positif	True Positive (TP)	False Negative (FN)
	Negatif	False Positive (FP)	True Negative (TN)

sumber: [33]

Dari tabel diatas dapat dijelaskan dengan keterangan dibawah

- **True Positive (TP):** Jumlah sampel yang benar-benar termasuk dalam kelas positif dan diprediksi dengan benar sebagai kelas positif oleh model.
- **True Negative (TN):** Jumlah sampel yang benar-benar termasuk dalam kelas negatif dan diprediksi dengan benar sebagai kelas negatif oleh model.
- **False Positive (FP):** Jumlah sampel yang sebenarnya termasuk dalam kelas negatif tetapi salah diprediksi sebagai kelas positif oleh model.
- **False Negative (FN):** Jumlah sampel yang sebenarnya termasuk dalam kelas positif tetapi salah diprediksi sebagai kelas negatif oleh model.

Dengan data tabel yang telah didapatkan, kemudian dapat dilakukan perhitungan akurasi, presisi, sensitivitas dan skor f-1 dengan rumus-rumus berikut.

1. **Akurasi (Accuracy):** Proporsi total prediksi yang benar dari semua prediksi yang dilakukan oleh model [32].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.8)$$

2. **Presisi (Precision):** Proporsi dari semua prediksi positif yang sebenarnya benar dibandingkan dengan total prediksi positif [33].

$$Precision = \frac{TP}{TP + FP} \quad (2.9)$$

3. **Recall (Sensitivitas):** Proporsi dari semua kelas positif yang berhasil diprediksi dengan benar dibandingkan dengan jumlah semua kelas positif yang sebenarnya [33].

$$Recall = \frac{TP}{TP + FN} \quad (2.10)$$

4. **F-1 Score:** Rata-rata harmonik dari presisi dan recall, memberikan keseimbangan antara keduanya [34].

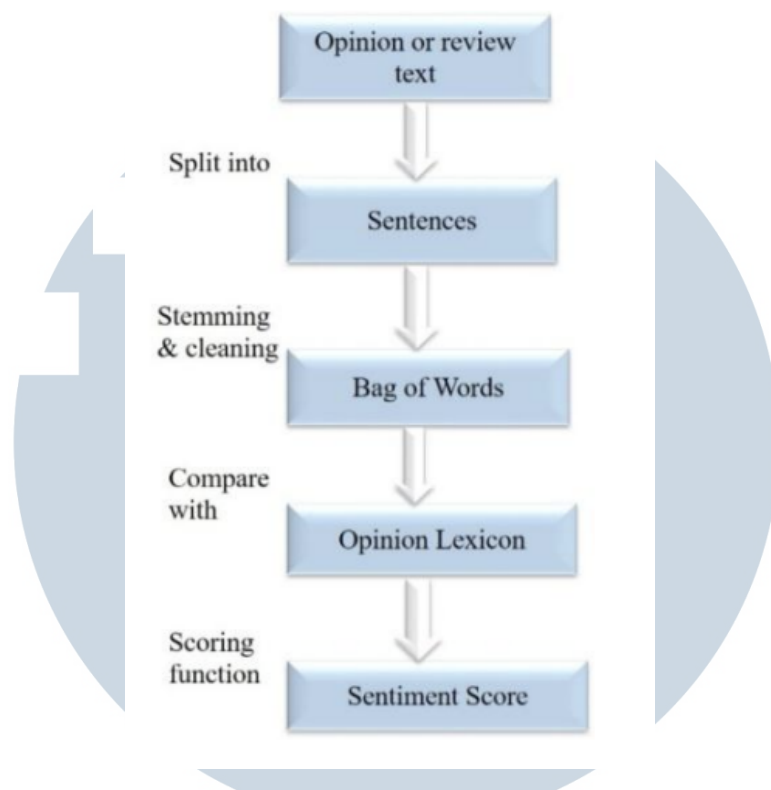
$$F1_Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.11)$$

2.8 Leksikon

Leksikon adalah kumpulan kata-kata atau istilah yang digunakan dalam suatu bahasa atau bidang khusus. leksikon mencakup semua kata yang dimengerti dan digunakan oleh penutur bahasa atau komunitas tertentu. Ini termasuk kata benda, kata kerja, kata sifat, frasa, dan ungkapan yang membentuk basis komunikasi dalam suatu bahasa. leksikon juga mencakup pemahaman tentang arti dan penggunaan kata-kata dalam konteks berbeda. Selain itu, leksikon juga dapat merujuk pada inventaris kata-kata dalam kamus atau sumber referensi yang mendokumentasikan kata-kata dalam sebuah bahasa dengan definisi dan penggunaan yang tepat. Dengan memahami leksikon suatu bahasa, seseorang dapat lebih efektif berkomunikasi dan memahami pesan yang disampaikan dalam konteks yang sesuai.[35]

Proses pelabelan leksikon dijelaskan pada gambar 2.3 Dasar Flowchart Kerja Leksikon. Pada penelitian "A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis" [36] melakukan komparasi pelabelan leksikon antara berbagai penyedia kamus leksikon. Penelitian melakukan proses dengan mengambil dataset, dataset kemudian di *preprocess* untuk menghasil kamus dan kata-kata, kamus kata-kata dari dataset kemudian akan dibandingkan dengan penyedia kamus leksikon dan perbandingan kata yang di hasilkan akan dilakukan pembobotan kalimat untuk mendapat nilai sentimen dari kalimat dataset.





Gambar 2.3. Dasar Flowchart Kerja Leksikon
sumber: [33]

UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA