

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Pembentukan Koperasi Desa Merah Putih menjadi salah satu program yang banyak diperbincangkan masyarakat sejak tahun 2025. Program ini bertujuan untuk memperkuat perekonomian rakyat melalui sistem koperasi yang berlandaskan prinsip kebersamaan dan kekeluargaan sebagaimana diatur dalam Undang-Undang Nomor 25 Tahun 1992 tentang Perkoperasian. Selain itu, penguatan koperasi juga didukung oleh Peraturan Pemerintah Nomor 7 Tahun 2021 tentang Kemudahan, Pelindungan, dan Pemberdayaan Koperasi dan UMKM serta Surat Edaran Menteri Koperasi dan UKM Nomor 1 Tahun 2025 tentang Petunjuk Pelaksanaan Pembentukan Koperasi Desa Merah Putih. Melalui kebijakan tersebut, pemerintah berharap koperasi mampu meningkatkan kesejahteraan masyarakat, memperluas akses terhadap kegiatan ekonomi produktif, serta mendorong kemandirian ekonomi yang berkelanjutan. Namun, pelaksanaan program ini juga menimbulkan berbagai tanggapan dari masyarakat, baik berupa dukungan, kritik, maupun keraguan terhadap efektivitasnya.

Berbagai tanggapan masyarakat terhadap pembentukan Koperasi Desa Merah Putih banyak disampaikan melalui media sosial, khususnya platform X. Platform ini memungkinkan pengguna menyampaikan opini secara terbuka dan cepat sehingga menjadi ruang diskusi digital yang aktif. Banyaknya komentar netizen mengenai Koperasi Desa Merah Putih menjadikan X sebagai sumber data yang relevan untuk memahami kecenderungan opini publik terhadap program

tersebut. Namun, jumlah komentar yang terus bertambah setiap hari menyebabkan proses pengamatan secara manual menjadi kurang efektif karena membutuhkan waktu dan tenaga yang besar[1].

Meningkatnya jumlah komentar dan tanggapan masyarakat di media sosial menyebabkan proses pengamatan opini publik secara manual menjadi kurang efektif karena membutuhkan waktu dan tenaga yang besar. Oleh karena itu, diperlukan suatu pendekatan yang mampu mengolah data teks secara lebih cepat dan terstruktur agar informasi yang terkandung di dalamnya dapat dianalisis dengan lebih efisien[2].

Salah satu pendekatan yang dapat digunakan adalah analisis sentimen, yaitu teknik untuk mengidentifikasi dan mengelompokkan opini ke dalam kategori positif, negatif, dan netral. Analisis sentimen mampu membantu proses pengolahan data teks secara sistematis sehingga kecenderungan pandangan masyarakat terhadap suatu isu dapat diketahui dengan lebih mudah. Dalam penelitian ini, analisis sentimen digunakan untuk menganalisis komentar netizen di media sosial X terkait pembentukan Koperasi Merah Putih agar diperoleh gambaran mengenai persepsi publik terhadap program tersebut[3].

Dalam mendukung proses klasifikasi sentimen yang lebih akurat, penelitian ini menggunakan model *IndoBERT* sebagai metode utama dalam pengolahan data teks. *IndoBERT* merupakan model berbasis *deep learning* yang dirancang khusus untuk memahami konteks bahasa Indonesia dengan lebih baik. Model ini memiliki kemampuan memahami hubungan antar kata dalam suatu kalimat sehingga dapat menghasilkan klasifikasi sentimen yang lebih akurat dibandingkan metode

konvensional. Penggunaan *IndoBERT* dinilai sesuai karena data yang digunakan berasal dari komentar netizen di media sosial yang umumnya menggunakan bahasa informal dan beragam[4].

Beberapa penelitian terdahulu menunjukkan bahwa analisis sentimen dapat dimanfaatkan untuk memahami opini publik di media sosial. Penelitian sebelumnya menjelaskan bahwa analisis sentimen pada data *Twitter* mampu memberikan gambaran mengenai pandangan masyarakat terhadap suatu kebijakan publik secara lebih cepat dan sistematis[5]. Penelitian lain menunjukkan bahwa teknik *Natural Language Processing* dapat mengolah data teks dalam jumlah besar sehingga pola opini masyarakat dapat diidentifikasi dengan lebih jelas[6]. Selain itu, penelitian mengenai penggunaan model *BERT* dalam analisis sentimen menunjukkan bahwa model berbasis *transformer* mampu meningkatkan akurasi klasifikasi karena dapat memahami konteks bahasa secara lebih baik[7].

Berdasarkan uraian tersebut, diperlukan suatu metode yang mampu mengolah data opini masyarakat secara otomatis dan akurat untuk mengetahui kecenderungan sentimen publik terhadap pembentukan Koperasi Desa Merah Putih. Oleh karena itu penelitian ini mengusulkan penggunaan model *IndoBERT* untuk melakukan analisis sentimen terhadap data yang diperoleh dari media sosial X. Hasil penelitian diharapkan mampu memberikan gambaran yang lebih objektif mengenai persepsi masyarakat terhadap program Koperasi Desa Merah Putih sekaligus menjadi referensi dalam pengambilan kebijakan di masa mendatang.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, maka rumusan

masalah dalam penelitian ini adalah bagaimana penerapan model *IndoBERT* dalam menganalisis sentimen publik terhadap pembentukan Koperasi Desa Merah Putih berdasarkan data dari media sosial *X*?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk menerapkan model *IndoBERT* dalam menganalisis sentimen publik terhadap pembentukan Koperasi Desa Merah Putih berdasarkan data dari media sosial *X*.

1.4 Batasan Masalah

Agar penelitian lebih terarah dan tidak melebar dari fokus kajian, maka penelitian ini dibatasi pada hal-hal berikut:

1. Data penelitian berupa unggahan pengguna berbahasa Indonesia yang berkaitan dengan topik Koperasi Desa Merah Putih dan diperoleh melalui proses *scraping* dari media sosial *X*.
2. Analisis sentimen dilakukan dengan mengklasifikasikan opini ke dalam tiga kategori yaitu positif, negatif, dan netral.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat akademis dan praktis, yakni sebagai referensi dalam kajian analisis data dan sentimen pada media sosial *X* serta menambah literatur ilmiah terkait opini publik. Penelitian ini dapat menggambarkan kecenderungan sentimen masyarakat terhadap pembentukan Koperasi Desa Merah Putih, menjadi bahan pertimbangan pihak terkait, dan membantu pengembangan metode analisis sentimen yang lebih efektif untuk penelitian selanjutnya.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini terdiri dari enam bab utama yang disusun secara sistematis sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini berisi uraian mengenai latar belakang penelitian, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, serta sistematika penulisan skripsi.

BAB 2 LANDASAN TEORI

Bab ini membahas teori-teori yang menjadi dasar dalam penelitian, meliputi konsep opini publik, perkembangan media sosial sebagai ruang komunikasi digital, analisis sentimen, serta teori mengenai koperasi dalam sistem perekonomian Indonesia. Selain itu, bab ini juga memuat tinjauan penelitian terdahulu yang relevan sebagai bahan perbandingan dan pendukung penelitian.

BAB 3 METODOLOGI PENELITIAN

Bab ini menjelaskan jenis dan pendekatan penelitian, sumber dan teknik pengumpulan data, tahapan penelitian, serta teknik analisis data yang digunakan untuk mengkaji sentimen publik terhadap pembentukan Koperasi Desa Merah Putih berdasarkan data dari media sosial X.

BAB 4 ANALISA DAN PERANCANGAN

Bab ini berisi penyajian hasil pengolahan data serta analisis sentimen publik terhadap pembentukan Koperasi Desa Merah Putih. Pada bab ini

juga dibahas kecenderungan opini masyarakat serta isu-isu utama yang muncul dalam diskusi publik.

BAB 5 IMPLEMENTASI DAN PENGUJIAN

Bab ini memaparkan proses pengolahan data secara lebih rinci, interpretasi hasil analisis, serta evaluasi terhadap temuan penelitian yang diperoleh dari data yang telah diklasifikasikan.

BAB 6 PENUTUP

Bab ini berisi kesimpulan dari seluruh rangkaian penelitian yang telah dilakukan serta saran yang dapat diberikan untuk pengembangan kebijakan maupun penelitian selanjutnya.

BAB 2

LANDASAN TEORI

2.1 *Artificial Intelligence (AI)*

Artificial Intelligence (AI) atau kecerdasan buatan merupakan cabang ilmu komputer yang bertujuan mengembangkan sistem yang mampu meniru kemampuan berpikir manusia, seperti mengenali pola, memahami bahasa, dan mengambil keputusan. Teknologi *AI* banyak digunakan untuk membantu pengolahan data secara cepat dan efisien[8].

Dalam analisis teks dan media sosial, *AI* digunakan untuk mengolah data berupa komentar, opini, dan tanggapan pengguna media sosial sehingga informasi yang terkandung di dalamnya dapat dipahami secara lebih sistematis. Pemanfaatan *AI* dalam analisis sentimen membantu proses klasifikasi opini ke dalam kategori positif, negatif, dan netral[9].

AI memiliki hubungan erat dengan *machine learning* dan *deep learning*. *Machine learning* merupakan bagian dari *AI* yang memungkinkan sistem mempelajari pola dari data, sedangkan *deep learning* merupakan pengembangan *machine learning* yang menggunakan jaringan saraf tiruan untuk memahami data yang lebih kompleks. Pada penelitian ini, *IndoBERT* termasuk model *deep learning* yang digunakan untuk menganalisis sentimen komentar netizen di media sosial X terkait pembentukan Koperasi Desa Merah Putih[10].

2.2 *Natural Language Processing (NLP)*

Natural Language Processing (NLP) merupakan cabang *Artificial Intelligence (AI)* yang berfokus pada kemampuan komputer dalam memahami,

mengolah, dan menganalisis bahasa manusia. *NLP* memungkinkan sistem mengenali makna dan konteks bahasa secara otomatis sehingga data teks dapat diproses dengan lebih sistematis[11].

Pada media sosial, *NLP* digunakan untuk menganalisis data teks seperti komentar, opini, dan tanggapan pengguna. Pemanfaatan *NLP* membantu proses pengolahan data dalam jumlah besar sehingga pola komunikasi dan opini masyarakat dapat diidentifikasi dengan lebih mudah[12].

Dalam analisis sentimen, *NLP* berperan dalam memahami makna suatu kalimat untuk menentukan kategori sentimen positif, negatif, dan netral. Oleh karena itu, *NLP* menjadi teknologi penting dalam penelitian analisis sentimen komentar netizen di media sosial X terkait pembentukan Koperasi Desa Merah Putih[13].

2.3 *IndoBERT*

IndoBERT merupakan model bahasa berbasis deep learning yang dikembangkan khusus untuk memproses dan memahami teks berbahasa Indonesia. Model ini adalah adaptasi dari arsitektur *Bidirectional Encoder Representations from Transformers (BERT)* yang telah dilatih menggunakan korpus teks berbahasa Indonesia, termasuk artikel berita, media sosial, dan dokumen daring. Dengan pelatihan tersebut, *IndoBERT* mampu menangkap struktur bahasa, kosakata, serta konteks linguistik yang umum digunakan dalam bahasa Indonesia, sehingga lebih efektif dibandingkan metode pemrosesan bahasa tradisional[14].

Secara teknis, *indoBERT* menggunakan arsitektur *transformer encoder* dengan mekanisme *self-attention* yang memungkinkan pemahaman hubungan antar

kata dalam kalimat secara *bidirectional*, yaitu dari kiri ke kanan dan sebaliknya. Pendekatan ini membuat setiap kata dianalisis berdasarkan konteks keseluruhan kalimat, menghasilkan representasi makna yang lebih akurat. Kemampuan ini memungkinkan *IndoBERT* digunakan untuk berbagai tugas *Natural Language Processing*, seperti klasifikasi dokumen, pengenalan entitas, dan analisis sentimen, termasuk pada teks media sosial yang sering memiliki variasi bahasa yang kompleks[15].

Dalam penelitian ini, *IndoBERT* diterapkan untuk analisis sentimen terhadap opini masyarakat di media sosial. Melalui proses *fine-tuning* dengan dataset yang relevan, model dilatih untuk mengklasifikasikan teks ke dalam kategori sentimen positif, negatif, atau netral. Pemanfaatan kemampuan pemahaman konteks bahasa ini diharapkan dapat meningkatkan akurasi analisis sentimen dan memberikan gambaran yang lebih *representatif* mengenai persepsi masyarakat terhadap isu yang diteliti[16].

2.4 *Fine-tuning Bidirectional Encoder Representations from Transformers (BERT)*

Fine-tuning merupakan proses lanjutan pada model *BERT* setelah tahap *pre-training* untuk menyesuaikan model dengan tugas tertentu menggunakan dataset yang lebih spesifik. Pada tahap ini, model yang telah memahami struktur dan konteks bahasa dilatih kembali agar mampu menghasilkan prediksi yang lebih akurat sesuai kebutuhan penelitian[17].

Dalam proses *fine-tuning*, data teks terlebih dahulu melalui tahap *tokenization* untuk mengubah kalimat menjadi token atau representasi numerik

yang dapat diproses oleh model. Selanjutnya, model dilatih menggunakan data yang telah memiliki label sehingga parameter model dapat diperbarui secara bertahap melalui proses optimasi. Untuk menghitung tingkat kesalahan prediksi model digunakan fungsi *cross-entropy loss* sebagai berikut:

$$L = - \sum_{i=1}^n y_i \log(\hat{y}_i) \quad (1)$$

Dimana L merupakan nilai loss, y_i adalah label sebenarnya, dan $\hat{y}_i$ merupakan hasil prediksi model. Nilai *loss* ini digunakan sebagai acuan untuk memperbaiki parameter model agar hasil prediksi semakin mendekati nilai yang sebenarnya. Sebelum menghitung loss, hasil keluaran model terlebih dahulu diubah menjadi bentuk probabilitas menggunakan fungsi softmax sebagai berikut:

$$P(y = i) = \frac{e^{z_i}}{\sum_{j=1}^k e^{z_j}} \quad (2)$$

Fungsi ini digunakan untuk menentukan kemungkinan suatu data termasuk ke dalam kelas tertentu, seperti sentimen positif, netral, atau negatif. Selanjutnya, model akan memperbarui parameter yang dimiliki agar kesalahan semakin kecil.

Pada penelitian analisis sentimen, *fine-tuning* digunakan untuk menyesuaikan kemampuan model *BERT* dalam memahami pola bahasa pada komentar media sosial. Melalui proses tersebut, model dapat mempelajari kecenderungan opini masyarakat berdasarkan data teks yang digunakan sehingga mampu meningkatkan akurasi klasifikasi sentimen terhadap komentar netizen di platform *X* terkait pembentukan Koperasi Merah Putih[18].

2.5 *Preprocessing Text*

Preprocessing text merupakan tahap awal dalam pengolahan data teks yang bertujuan untuk membersihkan dan menyiapkan data sebelum digunakan dalam proses analisis sentimen. Tahap ini dilakukan agar data yang diperoleh dari media sosial menjadi lebih terstruktur, konsisten, dan mudah dipahami oleh model. Data hasil scraping umumnya masih mengandung berbagai karakter yang tidak diperlukan, penggunaan bahasa tidak baku, serta data yang berulang sehingga perlu dilakukan proses preprocessing terlebih dahulu untuk meningkatkan kualitas dataset yang digunakan dalam penelitian[19].

Pada penelitian ini, tahapan preprocessing yang digunakan meliputi case folding, cleaning text, dan normalisasi kata slang. Case folding dilakukan dengan mengubah seluruh huruf pada teks menjadi huruf kecil (lowercase) untuk menyeragamkan bentuk penulisan kata. Selanjutnya, cleaning text dilakukan untuk menghapus karakter yang tidak diperlukan seperti URL, mention, hashtag, angka, tanda baca, emoji, dan simbol lainnya yang tidak memiliki pengaruh terhadap proses klasifikasi sentimen. Setelah itu, normalisasi kata slang dilakukan untuk mengubah kata tidak baku, singkatan, atau bahasa informal yang sering digunakan pada media sosial menjadi bentuk kata baku agar lebih mudah dipahami oleh model[20].

Selanjutnya, penelitian ini juga menerapkan penghapusan data duplikat, quality filtering, dan near duplicate removal. Penghapusan data duplikat bertujuan untuk menghilangkan data yang memiliki isi identik sehingga tidak terjadi pengulangan informasi dalam dataset. Quality filtering dilakukan untuk menyaring

data yang tidak relevan, terlalu pendek, kosong, atau memiliki kualitas teks yang rendah. Selanjutnya, near duplicate removal digunakan untuk menghapus data yang memiliki tingkat kemiripan sangat tinggi guna mengurangi bias dan meningkatkan kualitas data yang digunakan dalam proses pelatihan model[21].

Melalui serangkaian tahapan preprocessing tersebut, data komentar yang diperoleh dari media sosial X dapat diubah menjadi dataset yang lebih bersih, terstruktur, dan representatif. Dataset yang telah diproses selanjutnya digunakan sebagai masukan bagi model IndoBERT untuk melakukan klasifikasi sentimen ke dalam kategori positif, netral, dan negatif. Dengan kualitas data yang lebih baik, proses pelatihan model dapat berjalan lebih optimal sehingga menghasilkan performa klasifikasi yang lebih akurat[22].

2.6 *Confusion Matrix*

Confusion Matrix merupakan metode evaluasi yang digunakan untuk menilai kinerja model klasifikasi dengan membandingkan hasil prediksi model terhadap label data yang sebenarnya. Metode ini biasanya disajikan dalam bentuk tabel yang menunjukkan jumlah prediksi benar dan salah. Dalam klasifikasi, *confusion matrix* terdiri dari empat komponen utama, yaitu *True Positive (TP)*, *False Positive (FP)*, *True Negative (TN)*, dan *False Negative (FN)*, yang digunakan untuk menggambarkan bagaimana model mengklasifikasikan data yang dianalisis[23].

Pada permasalahan klasifikasi yang memiliki lebih dari dua kategori, digunakan konsep *multi class confusion matrix*. Pendekatan ini digunakan ketika model harus mengklasifikasikan beberapa kelas sekaligus, seperti pada analisis

sentimen yang terdiri dari sentimen positif, netral, dan negatif. Melalui tabel tersebut, peneliti dapat melihat distribusi hasil prediksi model pada setiap kelas serta mengetahui jumlah data yang berhasil diklasifikasikan dengan benar maupun yang mengalami kesalahan klasifikasi[24].

Berdasarkan nilai-nilai yang terdapat pada confusion matrix, performa model dapat diukur menggunakan beberapa metrik evaluasi. Metrik tersebut dihitung dengan memanfaatkan nilai *TP*, *TN*, *FP*, dan *FN* yang telah diperoleh dari hasil klasifikasi[25]. Salah satu ukuran yang digunakan adalah *accuracy*, *accuracy* digunakan untuk menunjukkan tingkat ketepatan model secara keseluruhan yaitu perbandingan antara jumlah prediksi yang benar dengan seluruh data yang diuji. Secara matematis yang menunjukkan tingkat ketepatan model secara keseluruhan, dengan rumus sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Selain itu, terdapat *precision* yang digunakan untuk mengetahui seberapa tepat hasil prediksi model pada suatu kategori tertentu. Dengan kata lain, *precision* menunjukkan seberapa banyak data yang diprediksi sebagai positif benar-benar sesuai dengan kondisi sebenarnya. Rumus *precision* adalah sebagai berikut:

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

Selanjutnya, *recall* digunakan untuk mengukur kemampuan model dalam menemukan seluruh data yang termasuk dalam suatu kategori negatif, positif, netral. Nilai *recall* menunjukkan seberapa banyak data yang seharusnya positif berhasil dikenali oleh model. Rumus *recall* dapat dituliskan sebagai berikut:

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

Untuk mendapatkan gambaran yang lebih seimbang antara *precision* dan *recall*, digunakan *F1-score*. Nilai ini merupakan gabungan dari kedua ukuran tersebut, sehingga dapat memberikan penilaian yang lebih menyeluruh terhadap kinerja model. Rumus *F1-score* adalah sebagai berikut:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

Dengan menggunakan keempat ukuran tersebut, peneliti dapat menilai kinerja model secara lebih lengkap, khususnya dalam melihat kemampuan model dalam mengelompokkan data ke dalam kategori sentimen positif, netral, dan negatif secara tepat.

2.7 Visual Studio Code

Visual Studio Code (VS Code) merupakan editor kode sumber (*source code editor*) yang dikembangkan oleh *Microsoft* dan banyak digunakan dalam pengembangan perangkat lunak karena memiliki fitur yang lengkap, ringan, serta mendukung berbagai bahasa pemrograman. *Visual Studio Code* menyediakan berbagai fasilitas seperti *syntax highlighting*, *IntelliSense*, *debugging*, *integrasi Git*, dan dukungan ekstensi yang membantu meningkatkan efisiensi pengembangan aplikasi. Kemudahan penggunaan dan kelengkapan fitur tersebut menjadikan *Visual Studio Code* sebagai salah satu editor kode yang banyak digunakan oleh mahasiswa maupun pengembang perangkat lunak[26].

Dalam penelitian ini, *Visual Studio Code* digunakan sebagai lingkungan pengembangan (*development environment*) untuk membangun aplikasi analisis

sentimen berbasis web menggunakan framework Streamlit dan bahasa pemrograman *Python*. Melalui *Visual Studio Code*, proses penulisan kode program, integrasi model *IndoBERT*, pengujian fungsi sistem, serta proses debugging dapat dilakukan secara lebih efektif dan terstruktur. Penggunaan *Visual Studio Code* dalam pengembangan aplikasi berbasis *Python* telah banyak diterapkan karena mampu mempermudah proses implementasi algoritma dan pengelolaan kode program secara lebih sistematis[27].

Selanjutnya, *Visual Studio Code* memiliki terminal terintegrasi dan dukungan berbagai ekstensi yang memudahkan proses instalasi pustaka (*library*), pengelolaan proyek, serta pengembangan aplikasi machine learning dan deep learning. Oleh karena itu, *Visual Studio Code* menjadi salah satu perangkat lunak pendukung yang berperan penting dalam proses pembangunan dan implementasi sistem analisis sentimen pada penelitian ini[19].

2.8 Python

Python merupakan bahasa pemrograman yang banyak digunakan dalam pengolahan data, pengembangan aplikasi, serta penelitian di bidang kecerdasan buatan. *Python* memiliki sintaks yang sederhana sehingga mudah dipahami dan digunakan dalam proses pengembangan program maupun analisis data[28].

Dalam bidang data science, *Python* digunakan untuk mengolah, menganalisis, dan memvisualisasikan data secara lebih efisien. *Python* didukung oleh berbagai *library* seperti *Pandas* dan *NumPy* yang membantu proses pengelolaan data, serta *Matplotlib* untuk visualisasi data[29].

Python juga banyak digunakan dalam bidang *Natural Language Processing (NLP)* dan *deep learning* karena mendukung berbagai library pengolahan teks dan pembelajaran mesin. Library seperti *Scikit-Learn*, *TensorFlow*, dan *Transformers* membantu proses analisis sentimen, klasifikasi teks, serta pengembangan model *deep learning* seperti *IndoBERT*. Oleh karena itu, *Python* menjadi salah satu bahasa pemrograman yang relevan dalam penelitian analisis sentimen pada media sosial X terkait pembentukan Koperasi Desa Merah Putih[30].

2.9 Google Collaboratory

Google Collaboratory merupakan layanan komputasi berbasis *cloud* yang disediakan oleh *Google* untuk menjalankan kode program secara daring melalui peramban *web*. Platform ini dirancang untuk memudahkan pengguna dalam menulis, menjalankan, serta mendokumentasikan kode program tanpa harus melakukan instalasi perangkat lunak tambahan pada komputer. Dengan memanfaatkan teknologi komputasi awan, pengguna dapat mengakses lingkungan pemrograman kapan saja dan dari berbagai perangkat yang terhubung dengan internet. Lingkungan kerja pada *Google Collaboratory* berbentuk *notebook* interaktif yang memungkinkan pengguna menggabungkan teks penjelasan, kode program, serta hasil eksekusi dalam satu dokumen yang terstruktur. Hal ini sangat membantu dalam proses dokumentasi penelitian karena setiap tahapan analisis dapat ditampilkan secara sistematis dan mudah dipahami[31].

Selain itu, *Google Collaboratory* juga menyediakan dukungan komputasi yang cukup tinggi, termasuk penggunaan *GPU* dan *TPU* yang dapat mempercepat proses pengolahan data serta pelatihan model berbasis *deep learning*. Fasilitas ini

menjadikan platform tersebut banyak dimanfaatkan dalam penelitian yang berkaitan dengan analisis data, pembelajaran mesin, serta pemrosesan bahasa alami. Dalam penelitian ini, *Google Collaboratory* digunakan sebagai lingkungan kerja utama untuk menjalankan berbagai tahapan analisis sentimen, mulai dari proses pembersihan data, pemrosesan teks, proses *encoding*, pelatihan model, hingga evaluasi performa model. Dengan dukungan komputasi yang tersedia serta kemudahan dalam pengelolaan kode dan dokumentasi, penggunaan *Google Collaboratory* membantu memperlancar proses penelitian dan memudahkan peneliti dalam mengelola seluruh tahapan analisis secara lebih efisien dan terorganisir[32].

2.10 Analisis Sentimen

Analisis sentimen merupakan teknik dalam *Natural Language Processing (NLP)* yang digunakan untuk mengidentifikasi dan mengelompokkan opini atau tanggapan dalam suatu teks ke dalam kategori tertentu. Analisis sentimen membantu proses pengolahan data teks secara otomatis sehingga informasi yang terkandung di dalamnya dapat dipahami dengan lebih mudah[33].

Tujuan analisis sentimen adalah untuk mengetahui kecenderungan pendapat terhadap suatu isu, produk, layanan, maupun kebijakan berdasarkan data teks yang dianalisis. Melalui analisis sentimen, data yang awalnya tidak terstruktur dapat diubah menjadi informasi yang lebih sistematis[34].

Secara umum, sentimen dibagi menjadi tiga kategori, yaitu sentimen positif, negatif, dan netral. Sentimen positif menunjukkan dukungan atau tanggapan baik, sentimen negatif menunjukkan kritik atau ketidakpuasan, sedangkan sentimen

netral menunjukkan pendapat yang bersifat biasa tanpa kecenderungan tertentu[35].

Pada media sosial, analisis sentimen banyak digunakan untuk menganalisis komentar dan tanggapan pengguna terhadap isu yang sedang berkembang. Pemanfaatan analisis sentimen membantu proses identifikasi kecenderungan sentimen masyarakat secara lebih cepat dan efisien, termasuk dalam penelitian mengenai komentar netizen di media sosial X terkait pembentukan Koperasi Desa Merah Putih[36].

2.11 Opini Publik

Opini publik merupakan pandangan, pendapat, atau tanggapan masyarakat terhadap suatu isu, kebijakan, maupun fenomena yang berkembang di lingkungan sosial. Opini publik dapat terbentuk melalui informasi, pengalaman, serta interaksi antar masyarakat terhadap suatu permasalahan tertentu[37].

Pembentukan opini publik dipengaruhi oleh berbagai faktor, seperti media informasi, lingkungan sosial, dan perkembangan isu yang sedang dibahas masyarakat. Perbedaan sudut pandang menyebabkan munculnya beragam opini, baik berupa dukungan, kritik, maupun tanggapan netral terhadap suatu isu[38].

Dalam perkembangan teknologi digital, media sosial menjadi salah satu sarana utama pembentukan opini publik. Platform media sosial memungkinkan masyarakat menyampaikan pendapat secara cepat dan terbuka sehingga berbagai tanggapan dapat tersebar luas dalam waktu singkat. Oleh karena itu, opini publik di media sosial menjadi sumber data yang penting dalam penelitian analisis sentimen terkait pembentukan Koperasi Desa Merah Putih di platform X[3].

2.12 Media Sosial

Media sosial merupakan platform digital yang memungkinkan pengguna untuk membuat, membagikan, dan bertukar informasi dalam bentuk teks, gambar, maupun video. Kehadiran media sosial mempermudah masyarakat dalam berkomunikasi dan menyampaikan informasi secara cepat kepada banyak pengguna. Selain sebagai sarana komunikasi, media sosial juga menjadi ruang penyampaian opini publik terhadap berbagai isu yang berkembang di masyarakat[39].

Karakteristik media sosial meliputi keterbukaan, kemudahan akses, serta kecepatan penyebaran informasi. Pengguna dapat memberikan komentar, tanggapan, maupun membagikan informasi secara langsung sehingga interaksi antar pengguna berlangsung secara aktif. Karakteristik tersebut menjadikan media sosial sebagai ruang digital yang dinamis dalam membentuk percakapan publik[40].

Dengan karakteristik tersebut, media sosial menjadi sumber data yang relevan dalam penelitian analisis sentimen. Berbagai komentar dan tanggapan pengguna dapat digunakan untuk memahami kecenderungan opini masyarakat terhadap suatu isu tertentu. Oleh karena itu, media sosial sering dimanfaatkan dalam penelitian untuk menganalisis sentimen publik, termasuk pada komentar netizen di platform X terkait pembentukan Koperasi Desa Merah Putih[41].

2.13 X

X merupakan salah satu platform media sosial yang memungkinkan pengguna menyampaikan informasi, opini, dan tanggapan secara cepat dalam

bentuk teks singkat. Platform ini banyak digunakan masyarakat untuk membahas berbagai isu sosial, politik, ekonomi, maupun kebijakan publik sehingga menghasilkan interaksi yang aktif antar pengguna[42].

Penelitian ini memilih platform *X* karena memiliki karakteristik percakapan yang terbuka dan mudah diakses sehingga memudahkan proses pengumpulan data komentar netizen. Selain itu, banyaknya tanggapan masyarakat terkait pembentukan Koperasi Desa Merah Putih di platform *X* menjadikan media sosial ini relevan sebagai sumber data penelitian analisis sentimen[43].

Penggunaan *X* dalam penelitian analisis sentimen juga didukung oleh banyaknya data teks yang dihasilkan pengguna setiap hari. Data tersebut dapat dimanfaatkan untuk memahami kecenderungan opini masyarakat, baik berupa sentimen positif, negatif, maupun netral terhadap suatu isu tertentu[12].

2.14 Koperasi Desa Merah Putih

Koperasi merupakan badan usaha yang beranggotakan orang perseorangan atau badan hukum koperasi yang menjalankan kegiatan berdasarkan prinsip kebersamaan dan kekeluargaan untuk meningkatkan kesejahteraan anggota dan masyarakat. Koperasi memiliki peran penting dalam mendukung kegiatan ekonomi masyarakat melalui pengelolaan usaha secara bersama[44].

Koperasi Desa Merah Putih merupakan program pembentukan koperasi yang bertujuan memperkuat perekonomian masyarakat melalui pengembangan koperasi di tingkat desa dan kelurahan. Program ini dibentuk untuk meningkatkan kemandirian ekonomi, memperluas akses kegiatan usaha produktif, serta mendukung pemberdayaan masyarakat secara berkelanjutan[45].

Pembentukan Koperasi Desa Merah Putih menimbulkan berbagai respons dari masyarakat. Sebagian masyarakat memberikan dukungan terhadap program tersebut karena dinilai mampu membantu penguatan ekonomi rakyat, sementara sebagian lainnya menyampaikan kritik maupun tanggapan terkait pelaksanaan program[46].

Berbagai respon masyarakat terhadap Koperasi Desa Merah Putih banyak disampaikan melalui media sosial, khususnya platform. Opini yang muncul di media sosial mencerminkan pandangan publik terhadap program tersebut sehingga dapat digunakan sebagai sumber data dalam penelitian analisis sentimen untuk memahami kecenderungan opini masyarakat[47].

2.15 Penelitian Terdahulu

Berikut adalah penelitian sebelumnya yang dijadikan acuan dan referensi dalam menyelesaikan masalah ini.

Tabel 2. 1: Penelitian Terdahulu

No	Nama	Tahun	Judul	Hasil
1.	A. Y. Pratama et al	2025	Analisis Sentimen Publik Terkait Danantara Menggunakan Algoritma IndoBERT pada Platform Media Sosial	Hasil penelitian menunjukkan bahwa model IndoBERT mampu memahami konteks bahasa Indonesia secara lebih baik sehingga menghasilkan klasifikasi sentimen yang lebih akurat pada data media sosial. Selain itu, proses fine-tuning membantu meningkatkan performa model dalam mengenali opini publik secara kontekstual[16].
2.	S. P. Mahardika et al.	2025	Analisis Sentimen Media Sosial	Penelitian membuktikan bahwa IndoBERT mampu memberikan hasil

			Menggunakan IndoBERT	klasifikasi yang lebih akurat dalam menganalisis opini masyarakat di media sosial karena dapat memahami konteks kalimat dan hubungan semantik antar kata. Kemampuan tersebut membantu model mengenali makna teks secara lebih mendalam sehingga proses identifikasi sentimen positif, negatif, dan netral menjadi lebih tepat[15].
3.	I. Mursidah et al.	2025	Klasifikasi Sentimen Google Play Store Aplikasi ChatGPT Berbahasa Indonesia Berbasis IndoBERT	Hasil penelitian menunjukkan bahwa IndoBERT efektif digunakan untuk klasifikasi sentimen bahasa Indonesia karena mampu mengenali pola bahasa informal, singkatan, serta variasi penulisan yang sering digunakan pada media sosial dan ulasan digital. Selain itu, kemampuan model dalam memahami konteks antar kata secara bidirectional membuat proses klasifikasi sentimen menjadi lebih akurat dibandingkan metode tradisional, sehingga mampu mengurangi kesalahan dalam memahami makna suatu kalimat[4].
4.	A. Wafda	2025	Aspect-Based Sentiment Analysis terhadap Cuitan Platform X tentang Kurikulum Merdeka	IndoBERT mampu mengidentifikasi sentimen positif, negatif, dan netral secara lebih tepat pada data media sosial karena model dapat memahami hubungan antar kata dalam

			Menggunakan IndoBERT	kalimat secara kontekstual. Kemampuan tersebut membuat model lebih efektif dalam mengenali makna kalimat, termasuk pada penggunaan bahasa informal, singkatan, maupun variasi bahasa yang sering ditemukan pada platform media sosial, sehingga hasil klasifikasi sentimen menjadi lebih akurat dan konsisten[35].
5.	P. IndoBERT	2025	Sentiment Analysis of Coretax: A Comparison of Manual, Transformers-Based, and Lexicon-Based Data Labeling on IndoBERT Performance	Hasil penelitian menunjukkan bahwa penggunaan IndoBERT dengan pendekatan transformer-based menghasilkan performa analisis sentimen yang lebih baik dibandingkan pendekatan tradisional. IndoBERT mampu memahami konteks bahasa Indonesia secara lebih mendalam sehingga hasil klasifikasi sentimen menjadi lebih akurat, terutama pada data media sosial yang menggunakan bahasa informal[2].
6.	T. C. M. Lovelindra	2025	Analisis Sentimen Mengenai Berita Pemberhentian Genosida Menggunakan Data X Dengan Algoritma BERT dan Naive Bayes Classifier	Penelitian menunjukkan bahwa model BERT mampu memahami pola bahasa dalam data media sosial sehingga menghasilkan klasifikasi sentimen yang lebih efektif dibandingkan pendekatan konvensional. Kemampuan model dalam memahami konteks antar kata membuat proses identifikasi sentimen positif, negatif, dan netral

				menjadi lebih akurat, terutama pada teks media sosial yang menggunakan bahasa tidak baku[34].
7.	D. K. Sumartha	2025	Opini Publik terhadap Kebijakan Kenaikan UKT Menggunakan Pendekatan IndoBERT	Hasil penelitian menunjukkan bahwa IndoBERT mampu mengurangi ambiguitas makna pada teks media sosial sehingga proses identifikasi opini publik menjadi lebih akurat dan sistematis. Kemampuan model dalam memahami konteks kalimat secara menyeluruh membuat klasifikasi sentimen positif, negatif, dan netral dapat dilakukan dengan lebih tepat[14].
8.	S. M. Anugerah et al.	2024	Sentiment Analysis Social Media for Disaster using Naïve Bayes and IndoBERT	Penelitian membuktikan bahwa kombinasi metode IndoBERT dan klasifikasi sentimen menghasilkan performa yang lebih baik dibandingkan metode machine learning biasa karena model mampu memahami konteks bahasa secara bidirectional. Kemampuan tersebut membuat proses identifikasi sentimen pada data media sosial menjadi lebih akurat, terutama pada teks yang menggunakan bahasa informal dan beragam[1].
9.	M. Khadapi dan V. M. Pakpahan	2024	Analisis Sentimen Berbasis Jaringan LSTM dan BERT terhadap Diskusi	Model BERT menunjukkan kemampuan yang baik dalam memahami konteks bahasa secara dua arah sehingga mampu meningkatkan akurasi klasifikasi sentimen terhadap data

				teks. Dengan kemampuan tersebut, model dapat mengenali hubungan antar kata dalam kalimat secara lebih tepat sehingga hasil analisis sentimen menjadi lebih efektif dan akurat[7].
10.	F. Koto dan T. Baldwin	2021	Pengembangan IndoBERT untuk NLP Bahasa Indonesia	Hasil penelitian menunjukkan bahwa IndoBERT sebagai model pretrained memiliki kemampuan yang sangat baik dalam berbagai tugas NLP, termasuk analisis sentimen. Model ini dilatih menggunakan korpus bahasa Indonesia dalam jumlah besar sehingga mampu memahami struktur bahasa dan variasi linguistik dengan lebih baik dibandingkan model umum[18].

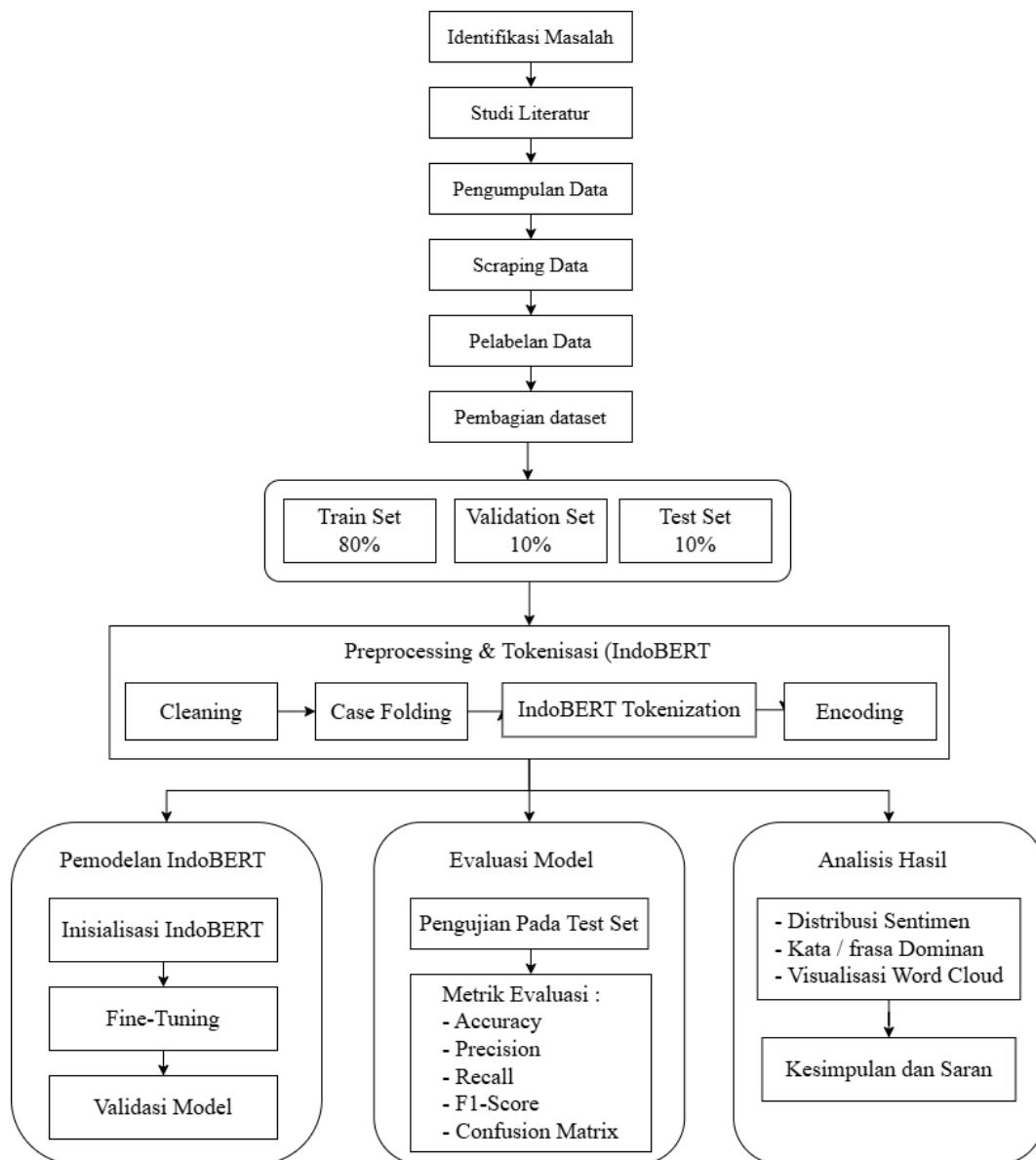
BAB 3

METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Tahapan penelitian merupakan langkah-langkah sistematis yang dilakukan untuk mencapai tujuan penelitian. Dalam penelitian ini, tahapan penelitian dilakukan untuk menganalisis sentimen opini masyarakat terhadap pembentukan Koperasi Desa Merah Putih melalui data yang diperoleh dari media sosial *X* dengan menggunakan metode *IndoBERT*. Proses penelitian dilakukan melalui beberapa tahapan utama yang meliputi pengumpulan data, pengolahan data teks, proses pelatihan model, hingga evaluasi hasil klasifikasi sentimen.

Tahapan penelitian ini dirancang untuk memastikan bahwa proses analisis sentimen dilakukan secara sistematis dan menghasilkan informasi yang akurat mengenai kecenderungan opini masyarakat terhadap kebijakan tersebut. Adapun tahapan lengkap dari penelitian ini dapat dilihat pada Gambar 3.1 berikut :



Gambar 3 . 1: Diagram Alur Tahapan Penelitian

3.2 Identifikasi Masalah

Dalam pengolahan data dari media sosial, salah satu kendala utama adalah banyaknya opini masyarakat yang tersebar dalam bentuk teks yang tidak terstruktur. Opini tersebut berisi berbagai tanggapan, baik berupa dukungan, kritik, maupun informasi biasa, sehingga sulit dipahami secara langsung tanpa proses analisis tertentu. Kondisi ini menjadi semakin kompleks karena bahasa yang digunakan di

media sosial cenderung tidak baku, sering menggunakan singkatan, serta memiliki makna yang bergantung pada konteks. Selain itu, pembentukan Koperasi Desa Merah Putih juga menimbulkan beragam respons dari masyarakat yang belum diketahui secara jelas kecenderungannya. Banyaknya data yang terus bertambah membuat analisis secara manual menjadi kurang efektif dan memakan waktu. Oleh karena itu, diperlukan suatu metode yang mampu mengolah data teks secara otomatis agar informasi yang terkandung di dalamnya dapat dipahami dengan lebih mudah.

Dalam hal ini, penggunaan model *IndoBERT* dianggap sesuai karena mampu memahami konteks bahasa Indonesia dengan lebih baik dibandingkan metode biasa. Dengan kemampuan tersebut, model ini dapat digunakan untuk mengelompokkan opini masyarakat ke dalam kategori sentimen positif, negatif, dan netral secara lebih akurat. Berdasarkan uraian tersebut, permasalahan dalam penelitian ini adalah bagaimana menerapkan model *IndoBERT* untuk menganalisis sentimen publik terhadap pembentukan Koperasi Merah Putih berdasarkan data dari media sosial X secara efektif dan efisien.

3.3 Studi Literatur

Studi literatur merupakan tahap awal yang dilakukan untuk mencari dan memahami berbagai sumber referensi yang berkaitan dengan penelitian ini. Sumber yang digunakan meliputi jurnal ilmiah, buku, serta artikel yang relevan dengan topik analisis sentimen dan perkembangan teknologi kecerdasan buatan, khususnya yang berkaitan dengan penggunaan model *IndoBERT*. Melalui kegiatan ini, peneliti

dapat memperoleh gambaran yang lebih jelas mengenai konsep serta pendekatan yang sesuai untuk digunakan dalam penelitian.

Selain itu, studi literatur juga membantu peneliti dalam memahami hasil-hasil penelitian sebelumnya yang memiliki keterkaitan dengan analisis sentimen pada media sosial. Dengan mempelajari berbagai penelitian terdahulu, peneliti dapat mengetahui metode yang sering digunakan, kelebihan, serta kekurangan dari setiap pendekatan. Hal ini menjadi dasar dalam menentukan metode yang paling tepat untuk menganalisis sentimen masyarakat terhadap pembentukan Koperasi Desa Merah Putih berdasarkan data dari media sosial X.

3.4 Pengumpulan Data

Pengumpulan data merupakan tahapan awal yang sangat penting dalam suatu penelitian, karena pada tahap ini peneliti memperoleh bahan utama yang akan dianalisis lebih lanjut. Data yang digunakan dalam penelitian ini bersumber dari media sosial X, yang memuat berbagai opini, tanggapan, serta pandangan masyarakat terkait pembentukan Koperasi Desa Merah Putih. Pemilihan media sosial X didasarkan pada sifatnya yang terbuka dan dinamis, sehingga memungkinkan peneliti untuk mengamati respons publik secara langsung terhadap suatu isu yang sedang berkembang.

Proses pengumpulan data dilakukan melalui teknik pengambilan data secara otomatis dari platform tersebut. Teknik ini memungkinkan peneliti memperoleh data dalam jumlah yang cukup besar dalam waktu yang relatif singkat. Data yang dikumpulkan berupa teks yang mencerminkan opini masyarakat, seperti komentar, cuitan, maupun tanggapan terhadap topik yang dibahas. Setiap data yang diperoleh

kemudian diseleksi agar tetap relevan dengan fokus penelitian, sehingga hanya informasi yang berkaitan dengan pembentukan Koperasi Desa Merah Putih yang digunakan dalam tahap analisis selanjutnya. Selain itu, dalam proses pengumpulan data, peneliti juga memperhatikan keaslian dan keberagaman sumber informasi. Hal ini dilakukan agar data yang digunakan dapat mewakili berbagai sudut pandang masyarakat, baik yang bersifat positif, negatif, maupun netral. Dengan demikian, data yang terkumpul diharapkan mampu memberikan gambaran yang lebih komprehensif mengenai persepsi publik terhadap isu yang diteliti.

Pengumpulan data merupakan tahapan awal yang sangat penting dalam penelitian karena pada tahap ini peneliti memperoleh bahan utama yang akan dianalisis lebih lanjut. Data yang digunakan dalam penelitian ini bersumber dari media sosial X yang memuat berbagai opini, tanggapan, dan pandangan masyarakat terkait pembentukan Koperasi Desa Merah Putih. Media sosial X dipilih karena memiliki sifat yang terbuka, dinamis, serta memungkinkan terjadinya interaksi publik secara langsung, sehingga peneliti dapat mengamati respons masyarakat terhadap isu yang sedang berkembang.

Data diperoleh dari media sosial X menggunakan teknik web scraping dengan kata kunci “Koperasi Desa Merah Putih”. Pengambilan data dilakukan pada periode 2025-2026, yang menghasilkan sebanyak 6.962 data awal berupa teks komentar pengguna. Teknik ini memungkinkan peneliti untuk mengumpulkan data dalam jumlah besar dalam waktu yang relatif singkat, sehingga proses pengumpulan data menjadi lebih efisien.

Data yang dikumpulkan berupa teks yang mencerminkan opini masyarakat, seperti cuitan, komentar, maupun tanggapan terhadap topik yang dibahas. Setiap data yang diperoleh kemudian diseleksi agar tetap relevan dengan fokus penelitian, sehingga hanya informasi yang berkaitan dengan pembentukan Koperasi Desa Merah Putih yang digunakan dalam tahap analisis selanjutnya.

Selanjutnya, dalam proses pengumpulan data, peneliti juga memperhatikan keaslian dan keberagaman sumber informasi. Hal ini dilakukan agar data yang digunakan dapat mewakili berbagai sudut pandang masyarakat, baik yang bersifat positif, negatif, maupun netral. Dengan demikian, data yang terkumpul diharapkan mampu memberikan gambaran yang lebih komprehensif mengenai persepsi publik terhadap isu yang diteliti.

3.5 *Scraping Data*

Pada penelitian ini, pengumpulan data dilakukan menggunakan teknik *scraping data*, yaitu proses pengambilan data secara otomatis dari media sosial X. Teknik ini digunakan untuk memperoleh data opini masyarakat dalam jumlah besar yang berkaitan dengan pembentukan Koperasi Desa Merah Putih. Pengambilan data dilakukan dengan menelusuri kata kunci yang relevan dengan topik penelitian, sehingga data yang diperoleh tetap sesuai dengan fokus kajian. Data yang dikumpulkan berupa teks unggahan yang mencerminkan tanggapan atau opini masyarakat terhadap pembentukan Koperasi Desa Merah Putih.

Melalui teknik ini, data dapat diperoleh secara lebih cepat dan beragam dibandingkan dengan pengumpulan secara manual, sehingga mampu memberikan gambaran yang lebih luas mengenai kecenderungan sentimen masyarakat. Dalam

pelaksanaannya, hanya data yang bersifat terbuka yang digunakan serta identitas pengguna tidak ditampilkan untuk menjaga privasi.

Pada penelitian ini, pengumpulan data dilakukan menggunakan teknik scraping data, yaitu proses pengambilan data secara otomatis dari media sosial X. Teknik ini digunakan untuk memperoleh data opini masyarakat dalam jumlah besar yang berkaitan dengan pembentukan Koperasi Desa Merah Putih. Pengambilan data dilakukan dengan menelusuri kata kunci yang relevan dengan topik penelitian sehingga data yang diperoleh tetap sesuai dengan fokus kajian. Data yang dikumpulkan berupa teks unggahan yang mencerminkan tanggapan atau opini masyarakat terhadap pembentukan Koperasi Desa Merah Putih.

Melalui teknik scraping ini, data dapat diperoleh secara lebih cepat dan beragam dibandingkan dengan pengumpulan data secara manual, sehingga mampu memberikan gambaran yang lebih luas mengenai kecenderungan sentimen masyarakat. Dalam pelaksanaannya, hanya data yang bersifat terbuka yang digunakan, serta identitas pengguna tidak ditampilkan untuk menjaga aspek privasi.

Data yang diperoleh melalui proses scraping masih mengandung data yang tidak relevan sehingga dilakukan tahap pra-pemrosesan data (data cleaning) yang meliputi beberapa tahapan, yaitu penghapusan bahasa non-Indonesia, penghapusan URL, mention, dan hashtag, penghapusan data duplikat, near duplicate removal dengan threshold kemiripan 0,85, serta quality filtering dengan menghapus teks yang terlalu pendek atau tidak bermakna.

Setelah seluruh tahapan tersebut dilakukan, diperoleh data bersih sebanyak

3.674 data yang kemudian digunakan sebagai dataset akhir dalam proses analisis penelitian.

3.6 Pelabelan Data

Pada tahap ini, data yang telah dikumpulkan selanjutnya diklasifikasikan berdasarkan kecenderungan sentimen yang terkandung di dalamnya. Setiap opini masyarakat yang diperoleh dari media sosial *X* dikelompokkan ke dalam tiga kategori utama, yaitu sentimen positif, netral, dan negatif. Pengelompokan ini bertujuan untuk memberikan gambaran yang lebih terarah mengenai bagaimana pandangan publik terhadap pembentukan Koperasi Desa Merah Putih. Opini yang berisi dukungan, harapan, atau penilaian baik dimasukkan ke dalam kategori positif, sedangkan pernyataan yang mengandung kritik, ketidaksetujuan, atau penilaian kurang baik dikategorikan sebagai negatif. Adapun pernyataan yang bersifat informatif tanpa menunjukkan kecenderungan sikap tertentu dikelompokkan sebagai netral.

Proses penentuan kategori dilakukan dengan menelaah isi setiap teks secara cermat agar makna yang terkandung di dalamnya dapat dipahami dengan baik. Ketelitian dalam tahap ini sangat penting karena hasil pengelompokan akan memengaruhi ketepatan analisis pada tahap berikutnya. Oleh karena itu, setiap data diperiksa secara seksama agar klasifikasi yang dihasilkan benar-benar mencerminkan maksud dari opini yang disampaikan. Dengan adanya proses ini, data yang semula beragam dapat disusun secara lebih terstruktur sehingga memudahkan peneliti dalam melihat pola sentimen masyarakat terhadap pembentukan Koperasi Desa Merah Putih secara lebih jelas dan menyeluruh.

Pada tahap ini, data yang telah dikumpulkan selanjutnya diklasifikasikan berdasarkan kecenderungan sentimen yang terkandung di dalamnya. Setiap opini masyarakat yang diperoleh dari media sosial X dikelompokkan ke dalam tiga kategori utama, yaitu sentimen positif, netral, dan negatif. Pengelompokan ini bertujuan untuk memberikan gambaran yang lebih terarah mengenai bagaimana pandangan publik terhadap pembentukan Koperasi Desa Merah Putih. Opini yang berisi dukungan, harapan, atau penilaian baik dimasukkan ke dalam kategori positif, sedangkan pernyataan yang mengandung kritik, ketidaksetujuan, atau penilaian kurang baik dikategorikan sebagai negatif. Adapun pernyataan yang bersifat informatif tanpa menunjukkan kecenderungan sikap tertentu dikelompokkan sebagai netral.

Proses penentuan kategori dilakukan melalui pelabelan menggunakan pendekatan *lexicon-based*, yaitu dengan memanfaatkan kamus kata positif dan negatif untuk mengidentifikasi kecenderungan sentimen pada setiap teks. Pendekatan ini memungkinkan proses klasifikasi awal dilakukan secara sistematis terhadap data dalam jumlah besar.

Meskipun demikian, hasil pelabelan ini bersifat pseudo-label, sehingga tidak sepenuhnya merepresentasikan kebenaran manual secara absolut, melainkan digunakan sebagai dasar awal dalam proses pembentukan dataset. Oleh karena itu, setiap data tetap ditinjau secara cermat untuk memastikan kesesuaian makna dan konteks opini yang disampaikan.

Dengan adanya proses ini, data yang semula bersifat tidak terstruktur dapat disusun menjadi lebih terorganisir sehingga memudahkan proses analisis lanjutan. Hasil pelabelan ini kemudian digunakan sebagai dataset pelatihan untuk model *IndoBERT* dalam tahap klasifikasi sentimen pada penelitian ini.

3.7 Pembagian Dataset

Setelah proses pelabelan selesai dilakukan, dataset selanjutnya dibagi menjadi tiga bagian yaitu *training set*, *validation set*, dan *testing set*. Pembagian dataset bertujuan agar proses pelatihan dan evaluasi model dapat dilakukan secara objektif serta mampu mengukur kemampuan model dalam mengklasifikasikan data yang belum pernah dilihat sebelumnya.

Pada penelitian ini, dataset dibagi dengan proporsi 80% untuk *training set*, 10% untuk *validation set*, dan 10% untuk *testing set*. *Training set* digunakan untuk melatih model *IndoBERT* agar mampu mempelajari pola sentimen yang terdapat pada data. *Validation set* digunakan untuk memantau performa model selama proses pelatihan dan membantu menentukan parameter terbaik. Sementara itu, *testing set* digunakan untuk mengukur performa akhir model setelah proses pelatihan selesai dilakukan.

Pembagian dataset dilakukan secara acak untuk memastikan setiap kelas sentimen memiliki distribusi yang seimbang pada masing-masing subset data. Dengan pembagian tersebut, model diharapkan mampu melakukan generalisasi dengan baik terhadap data baru yang tidak termasuk dalam proses pelatihan.

Setelah proses pelabelan selesai dilakukan, dataset selanjutnya dibagi menjadi tiga bagian, yaitu training set, validation set, dan testing set. Pembagian dataset ini bertujuan agar proses pelatihan dan evaluasi model dapat dilakukan secara objektif serta mampu mengukur kemampuan model dalam mengklasifikasikan data yang belum pernah dilihat sebelumnya.

Pada penelitian ini, dataset dibagi dengan proporsi 80% untuk *training set*, 10% untuk *validation set*, dan 10% untuk *testing set*. *Training set* digunakan untuk melatih model *IndoBERT* agar mampu mempelajari pola sentimen yang terdapat pada data. *Validation set* digunakan untuk memantau performa model selama proses pelatihan sekaligus membantu dalam penentuan parameter terbaik. Sementara itu, *testing set* digunakan untuk mengukur performa akhir model setelah proses pelatihan selesai dilakukan.

Pembagian dataset dilakukan menggunakan *stratified sampling* secara acak, sehingga distribusi setiap kelas sentimen (positif, netral, dan negatif) tetap seimbang pada masing-masing subset data. Dengan demikian, risiko bias data dapat diminimalkan dan model diharapkan mampu melakukan generalisasi dengan baik terhadap data baru yang tidak termasuk dalam proses pelatihan.

Hasil pembagian dataset adalah sebagai berikut: *training set* sebanyak 2.939 data (80%), *validation set* sebanyak 367 data (10%), dan *testing set* sebanyak 368 data (10%), dengan total keseluruhan data sebanyak 3.674 data.

3.8 *Preprocessing* dan Tokenisasi Data

Tahap *preprocessing* dan tokenisasi data dilakukan untuk menyiapkan data sebelum digunakan dalam proses pelatihan model *IndoBERT*. Data hasil pelabelan masih mengandung berbagai karakter yang tidak diperlukan sehingga perlu dibersihkan agar menghasilkan kualitas data yang lebih baik.

Tahapan *preprocessing* yang digunakan meliputi *cleaning text*, *case folding*, normalisasi kata slang, penghapusan data duplikat, *quality filtering*, dan *near duplicate removal*. *Cleaning text* dilakukan untuk menghapus URL, *mention*, *hashtag*, emoji, angka, dan simbol yang tidak diperlukan. *Case folding* digunakan untuk mengubah seluruh huruf menjadi huruf kecil. Normalisasi kata slang dilakukan untuk mengubah kata tidak baku menjadi kata baku, sedangkan *quality filtering* dan *near duplicate removal* digunakan untuk meningkatkan kualitas dataset.

Setelah *preprocessing* selesai dilakukan, data kemudian melalui tahap tokenisasi menggunakan *tokenizer IndoBERT*. Tokenisasi bertujuan mengubah teks menjadi token yang dapat dipahami oleh model. Hasil tokenisasi kemudian diubah menjadi representasi numerik berupa *input_ids* dan *attention_mask* melalui proses *encoding* sehingga dapat digunakan sebagai masukan pada model *IndoBERT*.

3.9 *Pemodelan IndoBERT*

Pada tahap ini dilakukan proses pembangunan model menggunakan *IndoBERT* yang telah dilatih sebelumnya pada korpus bahasa Indonesia. Model *IndoBERT* digunakan karena memiliki kemampuan memahami konteks bahasa Indonesia secara lebih baik dibandingkan metode konvensional.

Proses pemodelan diawali dengan inisialisasi model *IndoBERT* beserta *tokenizer* yang digunakan untuk mengubah data teks menjadi representasi numerik. Selanjutnya dilakukan proses *fine-tuning* menggunakan dataset *training* sehingga model dapat mempelajari pola sentimen positif, netral, dan negatif yang terdapat pada data penelitian.

Selama proses pelatihan, performa model dipantau menggunakan *validation* set untuk mengetahui kemampuan model dalam melakukan generalisasi terhadap data yang tidak digunakan dalam proses pelatihan. Tahap validasi juga digunakan untuk mengurangi risiko *overfitting* sehingga model yang dihasilkan memiliki performa yang lebih baik ketika digunakan pada data baru.

Pada tahap pemodelan ini dilakukan proses pembangunan model menggunakan *IndoBERT* yang telah dilatih sebelumnya pada korpus bahasa Indonesia. *IndoBERT* dipilih karena memiliki kemampuan yang lebih baik dalam memahami konteks bahasa Indonesia dibandingkan dengan metode pembelajaran mesin konvensional.

Proses pemodelan diawali dengan inisialisasi model *IndoBERT* beserta *tokenizer* yang digunakan untuk mengubah data teks menjadi representasi numerik yang dapat diproses oleh model. Selanjutnya dilakukan proses *fine-tuning* menggunakan dataset *training* sehingga model dapat mempelajari pola sentimen yang terdiri dari kategori positif, netral, dan negatif pada data penelitian.

Selama proses pelatihan, performa model dipantau menggunakan *validation* set untuk mengukur kemampuan generalisasi model terhadap data yang tidak

digunakan dalam pelatihan. Tahap validasi ini juga berfungsi untuk mengurangi risiko overfitting, sehingga model yang dihasilkan memiliki performa yang lebih stabil dan optimal ketika diterapkan pada data baru.

Serta parameter yang digunakan dalam penelitian ini adalah sebagai berikut: model *IndoBERT base*, jumlah *epoch* sebanyak 5, *learning rate* sebesar $2e-5$, *batch size* 16, *maximum sequence length* 128, *optimizer AdamW*, serta *random seed* 42 untuk memastikan reproduktibilitas hasil eksperimen.

3.10 Evaluasi Model

Evaluasi model dilakukan untuk mengetahui tingkat kinerja model *IndoBERT* dalam mengklasifikasikan sentimen. Evaluasi dilakukan menggunakan testing set yang tidak terlibat dalam proses pelatihan maupun validasi model. Metode evaluasi yang digunakan adalah *confusion matrix* yang menghasilkan informasi mengenai jumlah prediksi benar dan salah pada masing-masing kelas sentimen. Berdasarkan *confusion matrix* tersebut kemudian dihitung beberapa metrik evaluasi yaitu *accuracy*, *precision*, *recall*, dan *F1-score*.

Accuracy digunakan untuk mengukur tingkat ketepatan model secara keseluruhan. *Precision* digunakan untuk mengukur ketepatan prediksi pada setiap kelas sentimen. *Recall* digunakan untuk mengetahui kemampuan model dalam menemukan seluruh data yang termasuk ke dalam suatu kelas sentimen. Sementara itu, *F1-score* digunakan untuk memberikan ukuran performa yang mempertimbangkan keseimbangan antara *precision* dan *recall*.

Evaluasi model dilakukan untuk mengetahui tingkat kinerja model IndoBERT dalam mengklasifikasikan sentimen. Proses evaluasi menggunakan testing set yang tidak terlibat dalam proses pelatihan maupun validasi, sehingga dapat memberikan gambaran yang objektif mengenai kemampuan model dalam mengklasifikasikan data yang belum pernah dilihat sebelumnya.

Metode evaluasi yang digunakan adalah *confusion matrix*, yaitu teknik yang digunakan untuk menggambarkan jumlah prediksi benar dan salah pada setiap kelas sentimen. Berdasarkan hasil *confusion matrix* tersebut, kemudian dihitung beberapa *metrik evaluasi*, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*.

Accuracy digunakan untuk mengukur tingkat ketepatan model secara keseluruhan dalam melakukan klasifikasi. *Precision* digunakan untuk mengukur ketepatan prediksi pada setiap kelas sentimen. *Recall* digunakan untuk mengetahui kemampuan model dalam mendeteksi seluruh data yang termasuk ke dalam suatu kelas sentimen. Sementara itu, *F1-score* digunakan untuk memberikan ukuran performa yang mempertimbangkan keseimbangan antara *precision* dan *recall*.

Hasil evaluasi menunjukkan bahwa model memperoleh nilai *accuracy* sebesar 88,7%. Adapun nilai *precision*, *recall*, dan *F1-score* dihitung menggunakan *weighted average* untuk mempertimbangkan ketidakseimbangan distribusi data pada setiap kelas sentimen.

3.11 Analisis Hasil

Tahap analisis hasil dilakukan untuk memahami kecenderungan sentimen masyarakat terhadap pembentukan Koperasi Desa Merah Putih berdasarkan hasil

klasifikasi yang diperoleh dari model *IndoBERT*. Analisis dilakukan terhadap distribusi sentimen positif, netral, dan negatif yang muncul pada dataset penelitian. Selain melihat distribusi sentimen, analisis juga dilakukan terhadap kata dan frasa yang sering muncul pada masing-masing kategori sentimen. Analisis ini bertujuan untuk mengetahui topik atau isu utama yang banyak dibahas oleh masyarakat terkait pembentukan Koperasi Desa Merah Putih.

Hasil analisis kemudian divisualisasikan dalam bentuk grafik, diagram, dan *word cloud* untuk mempermudah interpretasi data. Melalui visualisasi tersebut dapat diketahui kecenderungan opini masyarakat serta kata-kata dominan yang sering muncul pada setiap kategori sentimen.

3.12 Kesimpulan dan Saran

Berdasarkan rangkaian penelitian yang telah dilakukan, dapat disimpulkan bahwa analisis sentimen terhadap opini masyarakat mengenai pembentukan Koperasi Desa Merah Putih dapat dilakukan dengan memanfaatkan data dari media sosial *X* menggunakan model *IndoBERT*. Melalui tahapan pengumpulan data, preprocessing teks, pelabelan data, hingga proses pelatihan dan evaluasi model, penelitian ini mampu mengklasifikasikan komentar masyarakat ke dalam tiga kategori sentimen, yaitu positif, netral, dan negatif. Penerapan model *IndoBERT* menunjukkan bahwa pendekatan berbasis *deep learning* mampu memahami konteks bahasa Indonesia dengan baik sehingga dapat digunakan untuk menggambarkan kecenderungan opini publik yang berkembang di ruang digital.

Sebagai saran, penelitian selanjutnya diharapkan dapat menggunakan jumlah data yang lebih banyak serta periode pengambilan data yang lebih panjang

agar hasil analisis sentimen dapat merepresentasikan opini masyarakat secara lebih luas. Selain itu, penelitian berikutnya juga dapat membandingkan model *IndoBERT* dengan metode atau model bahasa lain sehingga performa klasifikasi sentimen dapat ditingkatkan dan memberikan hasil analisis yang lebih akurat.