

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Masyarakat saat ini aktif menyampaikan opini mengenai berbagai isu pendidikan melalui media sosial, termasuk Kurikulum Merdeka. Platform X (Twitter) menjadi salah satu platform utama karena sifatnya yang cepat dan terbuka, sehingga menghasilkan banyak data tekstual [1]. Banyaknya opini publik ini menimbulkan tantangan dalam mengolah informasi secara manual, sehingga diperlukan metode analisis otomatis yang mampu menangkap sentimen dan topik utama secara efisien.

Penelitian sebelumnya menunjukkan bahwa model transformer seperti IndoBERT efektif dalam analisis teks berbahasa Indonesia berbasis aspek pada topik Kurikulum Merdeka, yang mendapat akurasi yang sangat tinggi pada klasifikasi aspek, namun performa klasifikasi sentimen masih berada pada tingkat yang lebih rendah. Hal ini mengindikasikan bahwa masih terdapat tantangan dalam mengklasifikasikan sentimen secara akurat, terutama pada data media sosial yang memiliki variasi ekspresi bahasa, seperti sarkasme dan ambiguitas [2]. Model BiLSTM memiliki kemampuan dalam menangkap ketergantungan jangka panjang pada teks, sehingga digunakan untuk mendukung representasi sekuensial yang dihasilkan oleh model transformer. Penelitian terdahulu belum banyak mengintegrasikan *Rule-Based* Heuristik untuk memperbaiki sentimen pasca-klasifikasi pada kasus ambiguitas dan sarkasme di data publik [3].

Urgensi penelitian ini muncul karena analisis sentimen saja tidak cukup untuk memahami persepsi publik secara menyeluruh, pemetaan topik juga diperlukan agar isu dominan terkait Kurikulum Merdeka dapat diidentifikasi. Selain itu, penting untuk menangani sarkasme dan ambiguitas opini yang sering muncul di media sosial, terutama di platform seperti Twitter. Sarkasme dapat mengaburkan makna asli dari sebuah opini, menyebabkan kesalahan dalam klasifikasi sentimen, seperti menganggap opini negatif yang disampaikan secara sarkastik sebagai

positif. Begitu juga dengan opini ambigu, dimana pernyataan yang mengandung makna ganda atau tidak jelas bisa disalahartikan jika hanya mengandalkan analisis sentimen dasar.

Untuk mengatasi permasalahan tersebut, penelitian ini memili pendekatan hybrid yang menggabungkan IndoRoBERTa sebagai ekstraktor fitur dan BiLSTM untuk sequence learning, serta *Rule-Based* Heuristik berbasis pola linguistik dan prioritas aturan untuk memperbaiki sentimen sarkasme dan ambigu pasca-klasifikasi. Selain itu, pemetaan topik dilakukan menggunakan metode *Keyword-Based Classifier* untuk pemetaan topik eksplisit dan *Non-Negative Matrix Factorization (NMF)* untuk mengidentifikasi topik laten, yang hasilnya divisualisasikan dalam dashboard aplikasi website berbasis Streamlit agar informasi lebih mudah diakses dan dianalisis secara interaktif.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana performa arsitektur Hybrid IndoRoBERTa-BiLSTM dalam mengklasifikasikan sentimen opini publik terkait Kurikulum Merdeka berdasarkan metrik accuracy, precision, recall dan F1-score?
2. Bagaimana peran algoritma *Rule-Based* Heuristik dalam memperbaiki model klasifikasi pada teks ambigu atau yang mengandung kata/frasa spesifik di tingkat implementasi sistem?
3. Topik-topik apa saja yang dapat diidentifikasi dalam diskursus Kurikulum Merdeka melalui kombinasi pendekatan manual berbasis *Keyword-Based Classifier* dan pendekatan *Non-Negative Matrix Factorization (NMF)*?

## 1.3 Batasan Masalah

Demi menjaga konsistensi arah penelitian, batasan masalah dalam penelitian ini ditetapkan sebagai berikut:

1. Objek penelitian adalah data tekstual opini publik yang diambil dari

platform X (Twitter) menggunakan teknik *crawling* dengan kata kunci "Kurikulum Merdeka".

2. Model Pre-trained yang digunakan sebagai lapisan pengekrasi fitur adalah *indonesian-roberta-base-sentiment-classifier* (IndoRoBERTa).
3. Proses pelabelan data menggunakan pendekatan pelabelan otomatis menggunakan model IndoRoBERTa.
4. Optimasi *Rule-Based* Heuristik dan pemetaan topik dengan strategi ganda (*Keyword-Based Classifier* dan *Non-Negative Matrix Factorization*) diimplementasikan pada lapisan aplikasi (Web Dashboard) sebagai fitur analisis pasca-klasifikasi, bukan diintegrasikan ke dalam arsitektur Neural Network saat pelatihan.
5. Visualisasi hasil akhir dibangun menggunakan kerangka kerja Streamlit.
6. Penelitian ini berfokus pada evaluasi performa model hybrid tunggal dan efektivitas koreksi pasca-klasifikasi, sehingga tidak melakukan komparasi performa dengan model deep learning lainnya secara terpisah.

#### 1.4 Tujuan Penelitian

1. Mengukur dan menganalisis kinerja arsitektur Hybrid IndoRoBERTa-BiLSTM dalam melakukan klasifikasi sentimen pada dataset Kurikulum Merdeka.
2. Mengimplementasikan dan menguji peran algoritma *Rule-Based* Heuristik sebagai mekanisme *safety net* dalam mengoreksi kesalahan klasifikasi model pada kasus ambiguitas dan sarkasme.
3. Mengidentifikasi dan memetakan topik-topik dominan dalam diskursus Kurikulum Merdeka untuk memberikan wawasan tematik.

#### 1.5 Manfaat Penelitian

1. **Manfaat Teoretis:** Penelitian ini diharapkan memberikan kontribusi pada pengembangan *Natural Language Processing (NLP)* untuk Bahasa Indonesia dengan menggabungkan model Hybrid IndoRoBERTa-BiLSTM

untuk klasifikasi sentimen. Pendekatan ini memperkaya pemahaman tentang penggabungan model Transformer dan RNN, serta mengaplikasikan *Rule-Based* Heuristik untuk menangani teks ambigu dan sarkasme. Selain itu, penggunaan NMF dan *Keyword-Based Classifier* untuk pemetaan topik membuka potensi baru dalam analisis diskursus.

2. **Manfaat Praktis:** Menyediakan purwarupa sistem pemantauan isu pendidikan berbasis data yang robust bagi Kemendikbudristek dan praktisi pendidikan untuk memahami aspirasi masyarakat secara cepat dan akurat.

## 1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun secara logis dan sistematis untuk memudahkan pemahaman terhadap keseluruhan isi penelitian. Adapun struktur penulisan terbagi menjadi lima bab, sebagai berikut:

### BAB I PENDAHULUAN

Bab ini berisi penjelasan mengenai latar belakang pentingnya analisis sentimen dan pemetaan topik terkait Kurikulum Merdeka, diikuti rumusan masalah, batasan masalah, tujuan, manfaat penelitian, serta sistematika penulisan yang menggambarkan arah dan ruang lingkup penelitian.

### BAB II TINJAUAN PUSTAKA

Bab ini memuat teori-teori yang mendasari penelitian, seperti konsep *Crawling* data, *preprocessing* teks, model IndoRoBERTa dan BiLSTM, *Random Over Sampling*, *heuristik*, serta metode pemetaan topik menggunakan *Keyword-Based Classifier* dan NMF. Tinjauan penelitian terdahulu dan kerangka berpikir juga dijelaskan untuk mendukung pengembangan sistem.

### BAB III METODE PENELITIAN

Bab ini menjelaskan langkah-langkah penelitian mulai dari pengumpulan data tweet, *preprocessing*, pelabelan otomatis, pembagian dan penyeimbangan data, hingga perancangan dan pelatihan model Hybrid IndoRoBERTa-BiLSTM. Selain itu dijelaskan penerapan *heuristik*, metode pemetaan topik, rancangan dashboard

Streamlit.

#### BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil pengolahan data, performa model hybrid, efektivitas *Rule-Based Heuristik* pada teks ambigu, serta hasil pemetaan topik. Bab ini juga menunjukkan implementasi sistem melalui dashboard Streamlit dan membahas temuan penelitian berdasarkan hasil analisis.

#### BAB V PENUTUP

Bab ini berisi kesimpulan mengenai efektivitas model, peran *Rule-Based Heuristik*, dan temuan topik, serta saran untuk pengembangan penelitian dan sistem pada tahap selanjutnya.

