

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pengeras suara merupakan salah satu alat yang digunakan untuk memberikan informasi kepada khalayak umum. Salah satu tempat dimana terdapat pengeras suara adalah masjid sebagai tempat ibadah bagi umat muslim. Dengan adanya pengeras suara (toa) di masjid maka suara adzan (peringat muslim untuk melaksanakan sholat) akan terdengar lebih jauh lagi. Selain untuk adzan, pengeras suara (toa) di masjid terkadang digunakan untuk memberikan informasi seperti berita kematian warga sekitar, tadarus al-qur'an, dan lain – lain [1]. Penggunaan pengeras suara yang berlebihan dapat menyebabkan beberapa pihak merasa terganggu terutama untuk masyarakat perkotaan yang multikultur sehingga bisa mengakibatkan terjadi konflik di masyarakat [2]. Dengan alasan tersebut pemerintah pun melalui Kementerian Agama Republik Indonesia mengeluarkan peraturan Surat Edaran Menteri Agama Nomor 05 Tahun 2022 tentang Pedoman Penggunaan Pengeras Suara di Masjid dan Mushala.

Dikutip dalam surat kabar *online* CNN Indonesia hadirnya aturan tersebut membuat riuh netizen Indonesia dimana ada masyarakat yang menyetujuinya dan adapula yang menolaknya dengan berbagai alasan [3]. Masyarakat pun mengeluarkan pendapat atau alasannya dengan berbagai cara, salah satunya dengan mengekspresikan lewat media social twitter [4]. Tak heran twitter menjadi salah satu tempat arus informasi yang cepat karena sifatnya *real-time* kemudian pengguna dari twitter juga berasal dari berbagai macam kalangan masyarakat [5].

Penggunaan twitter sebagai sarana masyarakat dalam mengekspresikan pendapatnya dapat diolah dengan menggunakan metode analisis sentimen untuk mengetahui bagaimana respon dari masyarakat terhadap aturan pengeras suara di masjid, apakah responnya positif, netral atau negatif. Metode analisis sentimen ini termasuk kedalam *text mining*. *Text mining* sendiri merupakan suatu proses untuk mengambil informasi dari data yang berbentuk teks [6]. Sementara itu, analisis

sentimen adalah pembelajaran komputasi yang menyangkut respon seseorang dalam bentuk pendapat, emosi dan sikap terhadap suatu peristiwa yang terjadi dimasyarakat [7].

IndoBERT adalah algoritma deep learning yang dikembangkan untuk kasus Natural Language Processing (NLP) khusus bahasa Indonesia. IndoBERT dibuat dari algoritma BERT. BERT kepanjangan dari Bidirectional Encoder Representations from Transformers, algoritma BERT ini dibuat dan digunakan oleh Google AI Language [8]. IndoBERT dapat digunakan untuk kasus NLP yaitu *Morpho-syntax and Sequence Labelling Tasks : Part-of-speech (POS) tagging, Named entity recognition (NER), and Dependency parsing, Semantic Tasks : Sentiment Analysis, and Summarization, Discourse Coherence Tasks : Next Tweet Prediction, and Tweet Ordering* [9].

Penelitian ini akan melakukan analisis sentimen peraturan penguat suara di masjid pada media social twitter menggunakan algoritma IndoBERT. Hasil akhirnya berupa pengelompokkan / klasifikasi apakah suatu cuitan di twitter termasuk tanggapan yang positif, netral atau negative.

1.2 Perumusan Masalah

1. Bagaimana menerapkan algoritma IndoBERT pada aplikasi analisis sentimen peraturan penguat suara di masjid ?
2. Bagaimana akurasi algoritma IndoBERT pada aplikasi analisis sentimen peraturan penguat suara di masjid ?

1.3 Tujuan

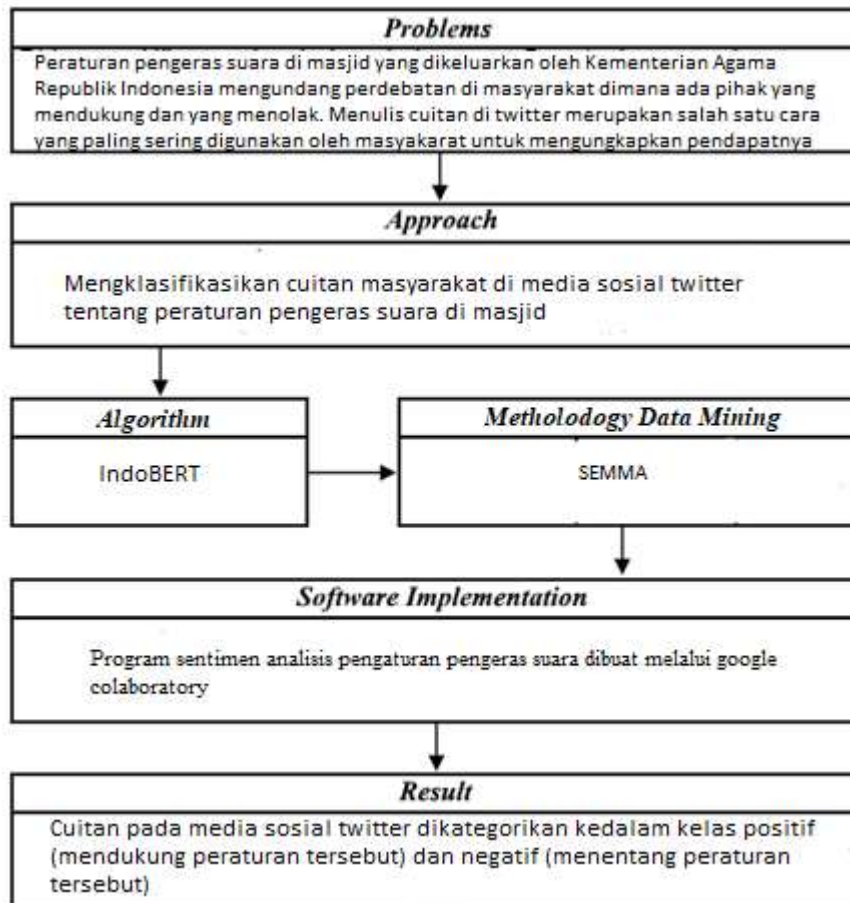
1. Menerapkan algoritma indoBert pada aplikasi analisis sentimen peraturan penguat suara di masjid.
2. Mengetahui kinerja algoritma indoBert pada aplikasi analisis sentimen peraturan penguat suara di masjid.

1.4 Batasan Masalah

1. Hasil yang diharapkan adalah algoritma indoBert dapat memprediksi suatu cuitan masuk kedalam kelas positif (setuju), netral, dan negatif (tidak setuju) terhadap peraturan pengeras suara di masjid.
2. Data berasal dari media sosial twitter dengan kata kunci aturan pengeras suara masjid OR aturan toa masjid.
3. Data diambil dengan menggunakan *library* twint dengan mengatur parameter *Limit* (batas jumlah cuitan yang ingin didapat) sebanyak 2000 dengan rentang waktu mulai dari 02-18-2022 sampai 03-25-2022.
4. Data dibagi menjadi 2 yaitu 80% sebagai data latih dan 20% sebagai data uji
5. Data yang digunakan hanya bahasa Indonesia
6. Pengembangan sistem menggunakan bahasa python
7. Semua program penelitian ini dilakukan menggunakan google colabatory
8. Metodologi penelitian menggunakan SEMMA dimana terdapat lima tahapan yaitu *sample* (pengambilan data), *explore* (mengeksplorasi data), *modify*, *model*, dan *assess* (pengujian)
9. Dilakukan data *preprocessing* untuk data cuitan sebelum melakukan pelatihan pada algoritma
10. Algoritma yang digunakan adalah IndoBERT.
11. IndoBERT yang digunakan adalah base model (phase-1 uncased) disingkat indobert-base-p1
12. Pengukuran kinerja model IndoBERT menggunakan confusion matrix berupa recall, precision, dan f1-score.

1.5 Kerangka Pemikiran

Adapun kerangka pemikiran dari penelitian tugas akhir ini direpresentasikan pada Gambar 1.1



Gambar 1.1 Kerangka pemikiran

Kerangka pemikiran di atas merupakan bentuk gambaran pemikiran yang akan diimplementasikan pada model analisis sentimen yang akan dibangun. Dengan mengangkat permasalahan adanya kebijakan pemerintah pusat tentang peraturan pengeras suara masjid. Selanjutnya dengan masalah tersebut bertujuan untuk membangun model analisis sentimen yang dapat mengklasifikasi komentar masyarakat tersebut dengan menggunakan metode *indoBERT* serta menggunakan metodologi SEMMA yang merupakan salah satu metode dalam *data mining*.

1.6 Metodologi Penelitian

Metodologi yang digunakan dalam penelitian ini adalah sebagai berikut:

1.6.1 Teknik Pengumpulan Data

Pada penelitian ini menggunakan dua teknik untuk mendapatkan data yaitu

1.6.1.1 Studi Literatur

Tahap awal melakukan pencarian informasi yang relevan dengan penelitian. Asal informasi penelitian ini didapat dari jurnal *online* (seperti skripsi), berita dari media yang kredibel seperti CNN indonesia ataupun buku/*ebook* yang membahas mengenai isu yang diangkat (aturan pengeras suara dimasjid), analisis sentimen, metode, algoritma yang dapat digunakan, proses untuk mendapatkan data, dan lain – lain.

1.6.1.2 Scraping

Untuk mendapatkan data dari twitter menggunakan teknik scraping. Scraping adalah salah satu teknik untuk mendapatkan data dari internet melalui web browser seperti chrome, firefox, safari, dan lain-lain. Web browser yang merupakan kumpulan kode html, css, javascript, dan data akan diekstrak untuk mendapatkan data yang diinginkan. Pada penelitian ini data yang diinginkan adalah cuitan pengguna twitter dan salah satu alat untuk scraping cuitan media sosial twitter adalah twint. Dengan menggunakan twint hanya perlu mengisi beberapa parameter diantaranya *Search* (kata kunci), *Limit* (batas cuitan yang ingin didapat), *Output* (*file* yang berisi data hasil scraping), dan lain-lain.

1.6.2 Metodologi Pengembangan

Penelitian ini menggunakan metodologi SEMMA. Metode SEMMA terdiri dari lima langkah yaitu

- Pengumpulan data (*sample*), bertujuan untuk mendapatkan data cuitan dari pengguna media sosial twitter yang mengandung kosakata aturan pengeras suara masjid atau aturan toa masjid.

- Eksplorasi data (*explore*), mengeksplor informasi dari data cuitan yang didapatkan. Beberapa informasi yang ingin didapat diantaranya *imbalace dataset*, duplikasi cuitan, kata yang sering muncul, dan lain-lain.
- Modifikasi data (*modify*), tahap ini dilakukan proses *preprocessing* pada cuitan yang didapatkan. Tahapan dalam *preprocessing* ini yaitu *feature selection* (menyeksi fitur atau kolom yang digunakan selama penelitian), dan data transformasi. Pada data transformasi dilakukan berbagai macam proses, proses tersebut diantaranya *case folding* (mengubah semua cuitan menjadi huruf kecil), menghapus baris baru, menghapus karakter/huruf yang berulang lebih dari 3 karakter/huruf berturut turut, menghapus kata yang tidak penting (*stop word*), mengubah kalimat menjadi kata (tokenisasi), dan mengganti kata-kata yang tidak baku menjadi baku.
- Model, pembuatan dan pelatihan model (indoBert). Algoritma indoBert didapat dari *library* transformers dengan tipe model yang dipilih yaitu phase-1 uncased. Proses pelatihan algoritma indoBert menggunakan google colaboratory dengan GPU sebagai *backend*-nya.
- Evaluasi model (*assess*). Model (algoritma indoBert) akan dievaluasi kinerjanya dalam memprediksi kelas suatu cuitan. Metode evaluasi yang digunakan adalah confusion matrix (mengukur akurasi).

1.7 Sistematika Penelitian

Laporan tugas akhir ini disusun dalam lima bab yaitu

BAB I PENDAHULUAN

Bab I tentang pendahuluan berisi Latar Belakang, Perumusan Masalah, Tujuan, Batasan Masalah, Kerangka Pemikiran, Metodologi Penelitian, dan Sistematika Penulisan

BAB II KAJIAN LITERATUR

Pada bab II mengenai kajian literatur memuat informasi mengenai *state of the art* (penelitian sebelumnya yang relevan dengan penelitian ini), dan landasan teori yang menjadi acuan untuk penelitian ini

BAB III METODOLOGI PENELITIAN

Bab III membahas analisis perangkat, analisis permasalahan dan metodologi SEMMA. Informasi yang diajabarkan yaitu spesifikasi minimal perangkat lunak dan keras, analisis permasalahan, pengumpulan data, eksplorasi data, modifikasi data yang mencakup data, pelatihan model, dan evaluasi model

BAB IV HASIL DAN PEMBAHASAN

Bab IV membahas implementasi dan pengujian sistem berupa hasil dari pengolahan data dan beberapa kode program.

BAB V SIMPULAN DAN SARAN

Bab kelima menjelaskan kesimpulan yang merupakan ringkasan dari penelitian dan saran untuk penelitian selanjutnya agar lebih baik.

