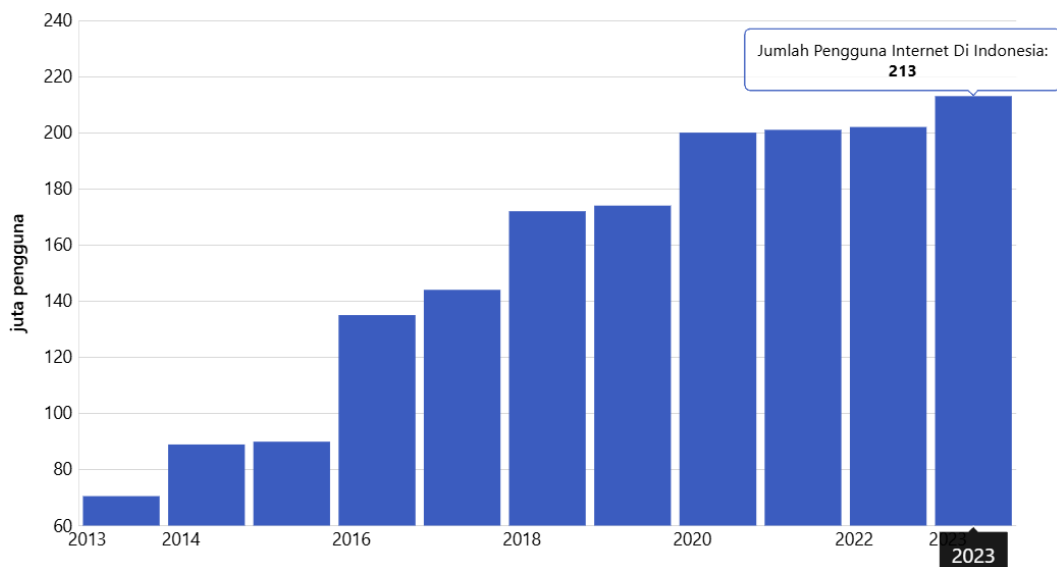


BAB I PENDAHULUAN

Pada bab pendahuluan ini akan dijelaskan terkait latar belakang permasalahan dari penelitian yang dilakukan, perumusan masalah spesifik yang ingin diselesaikan dan dibahas, tujuan dari dilakukannya penelitian, manfaat yang bisa diperoleh dari penelitian, serta bagaimana sistematika penelitian yang di implementasi pada laporan.

I.1 Latar Belakang

Pada zaman serba cepat, teknologi dan internet sudah menjadi suatu keharusan untuk dimiliki baik oleh individu maupun instansi. Perkembangan yang pesat ini tentunya mendorong dan menuntut seluruh pihak untuk andal dalam mengadopsi teknologi pada kegiatan sehari – hari. Data gambar I-1 menunjukkan perkembangan pengguna internet di Indonesia.



Gambar I-1 Perkembangan Pengguna Internet di Indonesia (Sumber : Katadata)

Salah satu contoh nyata yakni proses adopsi teknologi pada Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) yang juga merupakan Lembaga Pemerintahan Non Departemen (LPND) dengan tanggung jawab memberikan informasi terkait hal – hal seperti cuaca, iklim, maupun bencana di Indonesia. Dalam proses menyebarkan informasinya, BMKG mengeluarkan serta memaksimalkan berbagai *platform* baik *mobile application*, *website*, hingga media sosial dalam menyelesaikan tanggung jawabnya.

Feedback langsung dari pengguna aplikasi dapat dijadikan sebagai penggalan informasi apakah proses adopsi yang dilakukan sudah berjalan maksimal sesuai dengan yang diharapkan pengguna aplikasi. Selain itu optimalisasi aplikasi sangat penting untuk menyampaikan informasi yang akurat, khususnya untuk wilayah *Ring of Fire*. Wilayah ini merupakan wilayah yang memiliki banyak gunung aktif serta rentan terjadi bencana tsunami, gempa bumi, hingga berbagai bencana lain akibat dari aktivitas gunung-gunung tersebut. Sehingga untuk meningkatkan *awareness* masyarakat perlu adanya pengetahuan, peringatan, serta fasilitas mitigasi berbagai bencana yang memadai dan berkualitas tinggi. Dengan efektif dan efisien nya proses *transfer knowledge* tentunya akan berdampak positif juga terhadap kesadaran dan pengetahuan masyarakat terhadap rawan bencana di Indonesia. (Pradita, 2021).

Salah satu proses penilaian yang dapat dilakukan untuk mengetahui kualitas dari proses adopsi teknologi adalah dengan mengumpulkan *feedback* user yang dapat diakses via media sosial. Fakta lapangan pada sosial media BMKG menunjukkan masih adanya kekurangan yang dilihat dan dirasakan masyarakat. Selain itu masih banyak juga pertanyaan terkait informasi yang disampaikan sehingga menunjukan masih sulitnya proses pengolahan dan pemahaman informasi di masyarakat. Gambar I-2 menunjukkan beberapa temuan data terkait minimnya penyerapan informasi yang disediakan oleh BMKG.



Gambar I-2 Minimnya serapan informasi dari layanan yang tersedia

Data yang didapat dari lapangan ini tentunya perlu diolah lebih lanjut sebelum akhirnya dapat menjadi suatu *actionable insight*. Salah satu proses yang dapat dilakukan yakni *sentiment analysis* sehingga dapat mengetahui bagaimana

klasifikasi penilaian pengguna terhadap layanan tertentu. *Sentiment analysis* yang dilakukan pada penelitian ini dapat memberikan dampak positif berupa *guideline* terkait hal – hal yang masih perlu dilakukan peningkatan dan improvisasi.

Akan tetapi proses *sentiment analysis* ini memiliki beberapa tantangan seperti menangkap *semantic meaning* dan emosi berdasarkan komentar yang dilontarkan. Selain itu terdapat juga permasalahan teknis yang umum ditemui dalam perancangan pengolahan data dalam proses *sentiment analysis* seperti *handling* kata yang direpresentasikan dengan singkatan tertentu, bahasa - bahasa gaul, ataupun komentar yang tidak memiliki konteks tertentu seperti penggunaan *emoticons*. Hal ini membuat kompleksitas pengolahan data teks semakin tinggi dan tentunya dapat mempengaruhi akurasi dari analisa (Shuqin & Raga, 2024).

Sehingga dalam memaksimalkan proses *sentiment analysis*, penulis menggunakan konsep *aspect based* dengan memanfaatkan model LSTM sebagai *base model*, XLM-RoBERTa sebagai model untuk ekstraksi *embeddings* pada teks, serta menggunakan BERTopic dalam melakukan *subtopic* modeling pada *masing-masing* sentimen.

Alasan penulis menggunakan hybrid model LSTM + XLM-RoBERTa dalam penelitian *sentiment analysis* ini terinspirasi dari beberapa penelitian yang sudah dilakukan sebelumnya. Berdasarkan penelitian *sentiment analysis* yang dilakukan oleh (Tan dkk., 2022), menunjukkan performa *hybrid model* LSTM-RoBERTa memberikan akurasi 91,37 dibandingkan dengan LSTM tanpa *transfer learning* yang memiliki skor 77,56, model XGBoost yang hanya mendapatkan skor akurasi 73,84, dan model Naïve Bayes dengan skor akurasi 69,5. Hal ini membuktikan bahwa model LSTM yang dipadukan dengan model *pre-trained* merupakan model yang sesuai untuk tipe data teks khususnya komentar media sosial.

Pre-trained XLM-RoBERTa merupakan model yang sudah dilatih menggunakan berbagai bahasa berbeda – termasuk bahasa Indonesia – sehingga penulis dapat mempertahankan kondisi data teks dengan bahasa Indonesia. Tentunya hal ini akan memberikan sejumlah manfaat seperti proses penangkapan *semantic meaning* yang lebih mudah. (Santoso & Haryanto, 2023).

Goals dari penelitian ini tentunya memberikan *insight* berupa informasi terkait *aspect* apa saja yang perlu dipertahankan kualitasnya, serta bagian apa yang masih

perlu dilakukan peningkatan dalam pelayanannya. Gap dari penelitian terdahulu yakni penulis merancang model *multi-label* sebagai alat yang bisa dimanfaatkan untuk menarik *insight*. Selain itu penulis juga tidak hanya berfokus pada sentimen dari teks namun juga berfokus pada penentuan aspect dari suatu teks. Tentunya model LSTM yang dijadikan *base model* berdasarkan sumber yang penulis temukan bisa memberikan akurasi dan performa lebih baik dibandingkan algoritma *machine learning* (Khalaf dkk., 2024).

I.2 Perumusan Masalah

Rumusan masalah yang penulis angkat pada penelitian ini yakni:

- a) Bagaimana hasil model LSTM + XLM-RoBERTa terhadap kasus *sentiment analysis multi label*?
- b) Bagaimana hasil *subtopic* yang berhasil didapatkan dari permodelan BERTopic serta *mapping* terhadap ketiga *topic*?
- c) Bagaimana sebaran *sentiment* untuk masing-masing *topic* dan *subtopic*?

I.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang diambil, tujuan penelitian yang dilakukan adalah sebagai berikut:

- a) Mengetahui performa model LSTM + XLM-RoBERTa untuk kasus *sentiment analysis multilabel*.
- b) Mengetahui detail bahasan komentar dari masing-masing *topic* besar yang sering menjadi bahasan para pengguna khususnya pengguna yang menggunakan *platform* sosial media.
- c) Mengetahui bagaimana sebaran *sentiment* pada masing-masing *topic* untuk mendapatkan *actionable insight*.

I.4 Manfaat Penelitian

Manfaat dari penelitian yang dilakukan penulis klasifikasikan menjadi tiga bagian yakni:

- a) Dari sisi pihak lembaga BMKG, penelitian ini diharapkan dapat dijadikan acuan perbaikan serta improvisasi terhadap layanan yang ada. Tentunya dengan

mendapatkan acuan ini proses improvisasi akan lebih tepat sasaran serta dapat memberikan efektivitas dan efisiensi untuk waktu pengerjaan.

- b) Bagi pihak penulis selanjutnya, penelitian ini diharapkan bisa menjadi batu loncatan dalam menentukan apa yang dapat dilakukan selanjutnya. Hal ini dapat dilakukan dengan melakukan perbaikan dari kekurangan yang terdapat pada penelitian, mengingat banyaknya model baik *machine learning*, *deep learning*, maupun *pre-trained* model yang tersedia memberikan banyak peluang baru terkait objek penelitian.

I.5 Batasan Penelitian

Beberapa batasan penelitian yang penulis tetapkan antara lain:

- a) Akun yang dipakai sebagai objek penelitian merupakan @Stageofgowa dan @InfoBMKG.
- b) Menggunakan LSTM sebagai *base model* dalam melakukan *topic* dan *sentiment prediction*.
- c) Memanfaatkan model XLM-RoBERT-a untuk mendapatkan *embeddings* teks sebagai input pada *base model*.
- d) Platform sosial media yang digunakan yakni X dan Instagram dengan rentang waktu komentar dari tanggal 01 Januari 2015 s.d 22 Maret 2025.
- e) Menggunakan *pre-trained* model BERTTopic untuk mengetahui sub aspek mana yang masih memiliki banyak penilaian negatif dan positif dari masyarakat.

I.6 Sistematika Laporan

Berikut merupakan sistematika yang diimplementasikan penulis dalam merancang laporan:

- a) BAB I Pendahuluan

Bagian awal yang memberikan gambaran terkait latar belakang, detail rumusan masalah, tujuan, serta manfaat yang dapat dihasilkan dan didapatkan dari penelitian, serta bagaimana sistematika yang diimplementasikan pada laporan.

- b) BAB II Landasan Teori

Merupakan bagian yang dikhususkan untuk mengevaluasi teori – teori dan penelitian terdahulu yang relevan terkait model, kerangka standar, dan konsep umum terkait permasalahan masalah dan usulan solusi.

c) BAB III Metode Penyelesaian Masalah

Dalam bab ini diberikan gambaran langkah – langkah sistematis yang diambil dalam menyelesaikan permasalahan. Selain itu diberikan juga penjelasan umum terkait parameter yang digunakan untuk penilaian hasil, model yang digunakan, serta teknik analisa data yang semuanya ditampilkan menggunakan diagram alir.

d) BAB IV Penyelesaian Permasalahan

Pada bab ini diberikan penjelasan detail terkait proses pengumpulan dan pengolahan data baik dari *tools* yang dipakai, teknik pengolahan yang digunakan, penjelasan terkait implementasi dari masing-masing langkah, hingga proses analisis dalam pengambilan keputusan.

e) BAB V Validasi, Analisis dan Implikasi

Merupakan bagian yang membahas secara detail terkait validasi hasil olah data yang sudah dilakukan pada Bab IV. Hasil yang diberikan pada bab ini disesuaikan dengan masing-masing poin pada rumusan masalah yang diangkat.

f) BAB VI Kesimpulan dan Saran

Bagian akhir penelitian yang mengandung kesimpulan secara keseluruhan dari penelitian serta diberikan penjelasan atas jawaban dari berbagai rumusan masalah yang diangkat. Selain itu pada bab ini juga diberikan saran untuk penelitian selanjutnya serta kepada *stakeholder* terkait *actionable insight* yang perlu diimplementasikan guna menyelesaikan permasalahan yang ada.