

BAB 1 PENDAHULUAN

1.1. Latar Belakang

Di era yang sudah maju ini, pengguna aktif media sosial di seluruh dunia semakin pesat perkembangannya, menandakan peran media sosial yang besar untuk kebutuhan manusia di era ini. Di Indonesia, menurut We Are Social warga Indonesia rela menghabiskan waktu sekitar 188 menit per hari hanya untuk mengakses media sosial, yang mana data ini menunjukkan Indonesia berada di urutan negara ke-9 di seluruh dunia yang warganya gemar mengakses media sosial [1]. Di media sosial terdapat banyak aplikasi yang penggunaannya bisa berinteraksi secara *online*, seperti Facebook, YouTube, Instagram, dan lain-lain. Khususnya YouTube, sebuah platform media sosial yang dibuat untuk penggunaannya bisa berbagi informasi dalam bentuk video dan bisa mengomentari hasilnya [2]. Masih menurut We Are Social, di Indonesia penggunaannya bisa mencapai 143 juta, tingginya data ini tercermin dengan warga Indonesia menghabiskan 1.744 menit per bulan untuk mengakses YouTube [1].

Namun, dengan tinggi nya data pengguna pada YouTube di Indonesia bisa membawa dampak positif maupun negatif untuk para penggunaannya dalam menyampaikan komentar terhadap video, khususnya video yang berkaitan dengan penampilan Timnas Sepak Bola Senior di gelaran piala ASEAN 2024. Piala ASEAN merupakan ajang besar dalam kawasan Asia Tenggara. Di Indonesia, sepak bola cabang olahraga yang sangat populer. Data tahun 2024, Indonesia berada di urutan ke-15 di dunia negara paling menyukai sepak bola [3]. Sebagai negara penyuka sepak bola, banyak penonton yang mengharapkan lebih atas penampilan pemain serta strategi pelatih untuk meraih piala. Tetapi, tidak semua harapan itu menjadi kenyataan, pada turnamen Piala ASEAN 2024, Timnas Indonesia tidak lolos dari fase grup yang artinya tidak bisa melangkah lebih jauh dalam turnamen tersebut. Banyak penonton menyampaikan apa yang mereka rasakan saat menonton dalam bentuk mengomentari penampilan pemain atau strategi pelatih, ada yang memberikan semangat atau dukungan. Tetapi, tidak sedikit juga ber-komentar *toxic* hanya untuk mengekspresikan kekesalan atas apa yang mereka lihat, bisa berupa penghinaan, pelecehan dan rasisme [4].

Menurut penelitian jika terdampak komentar *toxic* di media sosial bisa membuat seseorang merasa menurun kepercayaan pada dirinya, cemas, depresi, hingga menimbulkan keinginan untuk bunuh diri [5]. Dalam konteks sepak bola akibat lainnya bisa berupa konflik suporter yang mana permasalahan tersebut bisa terbawa hingga dunia nyata. Dengan akibat yang separah itu harusnya pemerintah Indonesia serta pihak YouTube aktif dalam mencegah seperti memblokir akun atau menghapus komentar *toxic* khususnya berbahasa Indonesia untuk menciptakan suasana yang kondusif bagi semua pengguna di dalamnya.

Salah satu metode yang bisa dilakukan untuk mencegah hal tersebut terjadi antara lain menggunakan klasifikasi teks untuk mendeteksi komentar *toxic*. Banyak penelitian terdahulu yang sudah melakukan berbagai metode untuk mendeteksi komentar *toxic* seperti yang di lakukan oleh Andika et al., menggunakan metode CNN-LSTM dengan hasil F1-Score 84% dari YouTube dengan dataset berbahasa Indonesia [6]. Selain itu Nabiilah et al., menggunakan metode IndoBERT menghasilkan F1-Score 88% dengan dataset berbahasa indonesia dari instagram, kaskus dan X [7]. Park et al., juga melakukan dengan metode Random Forest menghasilkan F1-Score 94% dengan dataset bahasa Korea Selatan dari korpus berita [8].

Berbagai penelitian terdahulu telah dilakukan. Namun, tidak banyak studi yang melakukan penelitian dengan konteks deteksi komentar *toxic* pada penonton sepak bola timnas Indonesia dengan metode IndoBERT serta menggunakan teknik augmentasi dan teknik regularisasi. Untuk mengatasi kurangnya penelitian tersebut, penelitian ini dibuat. Penelitian ini menggunakan metode IndoBERT yang sudah dikembangkan untuk mendeteksi komentar *toxic* dan klasifikasi. Metode ini dipilih karena sudah dilatih dengan dataset berbahasa Indonesia.

1.2. Rumusan Masalah

Adapun rumusan masalah yang akan dibahas dalam penelitian ini:

1. Bagaimana hasil evaluasi IndoBERT dalam klasifikasikan komentar *toxic* bahasa Indonesia?
2. Bagaimana hasil IndoBERT dalam mendeteksi komentar *toxic* bahasa Indonesia?
3. Bagaimana perbedaan dampak penggunaan teknik augmentasi data dalam mempengaruhi hasil evaluasi *F1-Score* pada data *test*?
4. Bagaimana dampak perubahan nilai parameter dalam mempengaruhi hasil evaluasi *F1-Score* pada data *test*?
5. Bagaimana dampak teknik regularisasi dalam mengurangi *overfitting* pada IndoBERT?

1.3. Tujuan dan Manfaat

Penelitian ini dibuat untuk memberikan manfaat untuk para pembaca agar memberikan gambaran sejauh mana performansi metode IndoBERT dalam mengklasifikasikan dan deteksi komentar *toxic* dengan dataset bahasa Indonesia. Lalu, penelitian ini pun memiliki beberapa tujuan, seperti:

1. Mengetahui hasil evaluasi dari IndoBERT dalam klasifikasi komentar *toxic* bahasa Indonesia.
2. Mengetahui hasil deteksi komentar *toxic* oleh IndoBERT.
3. Mengetahui dampak penggunaan teknik augmentasi data pada IndoBERT.
4. Mengetahui dampak perubahan nilai parameter pada IndoBERT.
5. Mengetahui dampak penggunaan teknik regularisasi pada IndoBERT.

1.4. Batasan Masalah

Agar penelitian lebih terarah, diberikan beberapa batasan sebagai berikut:

1. Dataset dari penelitian ini difokuskan pada komentar bahasa Indonesia yang membahas performa pemain atau pelatih saat bertanding, lalu dilabeli secara manual sebagai *Toxic* dan *Non-Toxic*.
2. Komentar yang digunakan dalam deteksi dan klasifikasi hanya berbentuk teks.
3. Dataset diambil dari komentar yang berasal dari video cuplikan pertandingan dan *review* pertandingan di rentang waktu 8 Maret 2025-28 Mei 2025.
4. Pembagian data hanya data *training* dan *test*
5. Teknik pendukung hanya mencakup pra-pemrosesan data, augmentasi data, dan regularisasi.
6. Mengevaluasi model hanya menggunakan *Confussion Matrix* seperti *Accuracy*, *Precision*, *Recall* dan *F1-Score* dan hanya hasil evaluasi label *toxic* yang dianalisis.
7. Parameter yang digunakan hanya *Epoch*, *learning rate* dan *batch size*.

1.5. Metode Penelitian

Penelitian ini dilakukan dengan pendekatan studi literatur, analisis statistik, dan eksperiment komputasional. Penelitian dilakukan dengan mengumpulkan data berupa komentar pada YouTube, lalu masuk ketahapan pra-pemrosesan teks, Augmentasi Data (sinonim dan *random swap*), pembagian data, dan tokenisasi. Selanjutnya, tahap eksperimen dengan menganalisis berbagai eksperimen yang dicoba untuk melihat hasil evaluasi terutama akurasi dan F1-Score, terakhir mencoba simulasi untuk deteksi komentar *toxic* bahasa Indonesia.

1.6. Jadwal Pelaksanaan

Berikut merupakan rincian jadwal pelaksanaan yang dibuat untuk memastikan tugas akhir dapat berjalan dengan baik:

Tabel 1.1.1 Jadwal Pelaksanaan Tugas Akhir

No.	Deskripsi Tahapan	Bulan 1	Bulan 2	Bulan 3	Bulan 4	Bulan 5	Bulan 6
1	Studi Literatur						
2	Pengumpulan data dan pelabelan data secara manual						
3	Pra-pemrosesan Data						
4	Implementasi model IndoBERT						
5	Eksperimen teknik dan parameter pada model						
6	Evaluasi dan mencatat hasil performansi model						
7	Penyusunan Laporan TA						
8	Revisi dan finalisasi Laporan TA						
9	Sidang TA						