

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Informasi dan komunikasi bergerak sangat cepat seiring dengan perkembangan zaman, dan media sosial telah menjadi alat komunikasi yang umum dan penting di dunia modern saat ini. Menurut penelitian yang dilakukan oleh (Kemp., 2021), lebih dari empat miliar orang di dunia menggunakan jejaring sosial untuk berkomunikasi. Jelas bahwa aplikasi media sosial memberikan berbagai keuntungan kepada penggunanya. Namun, itu juga bisa digunakan untuk kejahatan. Salah satu target jahatnya adalah *Cyberbullying*.

*Cyberbullying* didefinisikan sebagai tindakan seseorang atau sekelompok orang yang menggunakan teknologi digital (Slonje and Smith, 2008). *Cyberbullying* memiliki efek yang lebih kuat dan tahan lama dibandingkan konsep klasik *bullying* karena dapat menjangkau banyak orang dengan cepat. Selain itu, menghapus dokumen online yang berisi konten berbahaya tersebut memungkinkan memakan waktu lama atau dalam beberapa kasus tidak mungkin. Meskipun *cyberbullying* tampaknya tidak menyebabkan kerugian fisik langsung pada korbannya, namun berpotensi menyebabkan gangguan psikologis seperti depresi, kehilangan kepercayaan diri, kurang fokus dan bahkan upaya bunuh diri (Mattson *et al.*, 2015). Kehadiran media sosial mempermudah pemakai teknologi untuk melakukan aksi *cyberbullying*, pelaku dapat memposting tulisan kejam atau mengunggah foto yang berhubungan dengan pribadi yang bertujuan untuk mengintimidasi dan merusak reputasi korban sehingga korban merasa malu, kecewa dan tersakiti, sedangkan pelaku merasa bahagia dan puas karena tujuannya telah tercapai (Siwi, Utami and Baiti, 2018). Twitter adalah sebuah platform media sosial yang sering digunakan oleh masyarakat untuk berkomunikasi dan menjalin relasi.

Pengguna Twitter bukan hanya pengguna individu perseorangan. Twitter juga digunakan oleh kelompok tertentu seperti komunitas, Lembaga negara, hingga toko berbasis daring. Tujuan pemakaian juga bermacam - macam mulai dari promosi, berbagi informasi tentang kinerja/hal yang telah dilakukan, bahkan mendapatkan informasi pribadi secara ilegal. tetapi tidak semua pemakai media sosial, menggunakan

teknologi ini dengan bijaksana, dan banyak pengguna menggunakan Twitter untuk melakukan tindakan jahat seperti aksi penipuan, penyebaran berita *hoax*, menulis hal-hal yang mengandung ujaran kebencian seperti perundungan secara daring (*cyberbullying*). Hal ini merupakan dampak negatif adanya platform media sosial di kehidupan masyarakat. Keadaan yang membuat klasifikasi *cyberbullying* sangat sulit ditangani, dikarenakan tidak ada standar yang benar-benar resmi untuk ujaran kebencian seperti rasisme, seksisme, toksisitas (Golem, Karan and Šnajder, 2018). Sementara beberapa pemakai media Twitter mengalami cuitan/komentar seseorang yang menyakitkan, namun mungkin bagi orang lain, hal tersebut bukan masalah, banyak korban yang mengalami ujaran kebencian tidak melapor ke pihak berwajib, mungkin karena takut atau merasa tidak dipedulikan ketika melapor, dan hal tersebut membuat perilaku ujaran kebencian akan selalu tumbuh subur dan berkembang pesat

Berbagai macam penelitian terkait dengan *cyberbullying* telah dilakukan. Sebagai contohnya salah satu penelitian pada komentar twitter telah dilakukan oleh Habib Faizal Fadli, Ahmad Fathan Hidayatullah identifikasi komentar *cyberbullying* menggunakan Metode LSTM dan BiLSTM, kelas target terdiri dari 2 kelas yaitu kelas *cyberbullying* dan *non-cyberbullying* dengan 6835 sampel. Masing-masing Akurasi dari algoritma tertera sebagai berikut LSTM 93% dan BiLSTM 95% Lalu, untuk nilai dari F1-Score dari masing-masing algoritma sebagai berikut LSTM 92% dan BiLSTM 93%.

Penelitian mengenai identifikasi ujaran kebencian pada media sosial twitter telah dilakukan oleh Aditya Perwira Joan Dwitama, Syarif Hidayat menggunakan *deep learning* dengan metode *Convolutional Neural Network*. Memiliki 12 kelas dengan 131.169 sampel data dan akurasi yang dihasilkan sebesar 98,76%.

Berdasarkan pembahasan tersebut, penelitian ini akan melakukan pembuatan model *deep learning* untuk membandingkan 3 metode yaitu LSTM, BiLSTM dan CNN untuk melakukan klasifikasi *text cyberbullying*. Penelitian ini diharapkan dapat memberikan model terbaik dan performa yang baik dalam melakukan klasifikasi teks *cyberbullying* dan dapat membantu pengguna maupun pihak aparat keamanan siber dalam melakukan pencegahan tindakan *cyberbullying*.

Tinjauan Agama Islam juga dilakukan untuk melihat dasar hukum dari

*cyberbullying*. Hal ini sebagaimana di jelaskan dalam surat

يَا أَيُّهَا الَّذِينَ آمَنُوا لَا يَسْخَرْ قَوْمٌ مِّنْ قَوْمٍ عَسَىٰ أَنْ يَكُونُوا خَيْرًا مِّنْهُمْ وَلَا نِسَاءٌ  
 مِّنْ نِّسَاءٍ عَسَىٰ أَنْ يَكُنَّ خَيْرًا مِّنْهُنَّ وَلَا تَلْمِزُوا أَنْفُسَكُمْ وَلَا تَنَابَزُوا بِالْأَلْقَابِ  
 بِئْسَ الْإِسْمُ الْفُسُوقُ بَعْدَ الْإِيمَانِ وَمَنْ لَّمْ يَتُبْ فَأُولَٰئِكَ هُمُ الظَّالِمُونَ

Artinya: “Hai orang-orang yang beriman, janganlah sekumpulan orang laki-laki merendahkan kumpulan yang lain, boleh jadi yang ditertawakan itu lebih baik dari mereka. Dan jangan pula sekumpulan perempuan merendahkan kumpulan lainnya, boleh jadi yang direndahkan itu lebih baik. Dan janganlah suka mencela dirimu sendiri dan jangan memanggil dengan gelaran yang mengandung ejekan. Seburuk-buruk panggilan adalah (panggilan) yang buruk sesudah iman dan barangsiapa yang tidak bertobat, maka mereka itulah orang-orang yang zalim.”

## 1.2 Perumusan Masalah

Berdasarkan latar belakang di atas, maka permasalahan penelitian ini dapat diidentifikasi antara lain sebagai berikut:

1. Bagaimana implementasi dari hasil perbandingan 3 model *deep learning* untuk klasifikasi komentar *cyberbullying* pada komentar twitter.
2. Bagaimana tingkat akurasi yang diberikan dari hasil perbandingan 3 model untuk klasifikasi komentar *cyberbullying* di twitter.
3. Bagaimana sudut pandang Islam mengenai *cyberbullying*.

## 1.3 Tujuan Penelitian

1. Melakukan perbandingan 3 metode *deep learning* terhadap klasifikasi komentar *cyberbullying* menggunakan CNN, LSTM dan BiLSTM.
2. Melakukan evaluasi performa model algoritma CNN, LSTM dan BiLSTM.
3. Mengetahui dan memahami sudut pandang Islam mengenai *cyberbullying* serta implementasi dari penerapan metode *deep learning*.

## 1.4 Manfaat Penelitian

1. Model Klasifikasi yang dihasilkan pada penelitian ini diharapkan dapat membantu Badan Siber Nasional (Basinas) untuk mendeteksi konten

*cyberbullying*.

2. Memberi pengetahuan bagi pengguna media sosial khususnya pengguna twitter untuk lebih bijaksana dalam memberikan komentar.
3. Memberikan pengetahuan *cyberbullying* mengenai sudut pandang Islam

### **1.5 Batasan Penelitian**

Adapun batasan masalah dari permasalahan yang ada, yaitu :

1. Sumber data yang digunakan berasal dari repositori Mendeley.
2. Komentar twitter yang diklasifikasi adalah komentar bahasa Inggris.
3. Metode yang digunakan yaitu metode CNN, LSTM dan biLSTM.
4. Bahasa pemrograman yang digunakan adalah Python dengan menggunakan *framework* Tensorflow dan *library* Keras.
5. Dataset yang digunakan berformat .csv