

## 1. Pendahuluan

Media sosial telah menjadi alat komunikasi dan interaksi yang paling banyak digunakan setiap orang selama beberapa tahun terakhir. Interaksi langsung antar manusia semakin berkurang karena orang-orang cenderung berkomunikasi secara tidak langsung melalui *smartphone*. Dengan demikian, cukup sulit untuk mengenali kepribadian seseorang. Namun, apa yang tertulis di media sosial dapat membantu kita mendapatkan informasi yang dibutuhkan karena orang-orang menghabiskan banyak waktu untuk memeriksa media sosial dan mengekspresikan perasaan dan pikiran mereka melalui status, komentar, dan *update* [1].

Kepribadian adalah pola pikir setiap individu yang mendefinisikan perilaku, perasaan, sikap dan karakteristik sebagai perbedaan sifat setiap manusia. Bagaimana seseorang dapat berbicara dan menulis dapat merefleksikan kepribadiannya [2]. Salah satu pendekatan yang dapat digunakan untuk memahami kepribadian adalah pendekatan Big Five Personality Traits yang juga dikenal sebagai OCEAN model. Model ini dikembangkan oleh Paul T. Costa, Jr. dan Robert R. McCrae. Model dimensi kepribadian "*Big Five*" telah muncul sebagai salah satu ukuran struktur kepribadian yang paling banyak diteliti dan dianggap baik dalam beberapa tahun terakhir [3]. Sifat kepribadian yang ada pada model tersebut terdiri dari 5 dimensi kunci yaitu *Openness, Conscientiousness, Extraversion, Agreeableness dan Neuroticism* [4]. Dengan menggunakan metode *Big Five* tersebut, kepribadian seseorang dapat dideteksi pada sosial media. Pada penelitian ini, media sosial Twitter dipilih untuk mengimplementasikan metode *Big Five* untuk dapat mendeteksi kepribadian seseorang berdasarkan berbagai cuitan pengguna [5].

Untuk menganalisis setiap cuitan pengguna yang terkait dengan kepribadian tentunya dapat dilakukan secara manual. Namun cara tersebut kurang efisien karena membutuhkan waktu yang sangat lama untuk dapat menganalisis kepribadian berdasarkan cuitan setiap pengguna. Selain itu, banyaknya cuitan pengguna dan pengguna yang bervariasi sehingga tidak memungkinkan untuk dapat dianalisis secara manual. Salah satu teknologi yang diimplementasikan untuk mendeteksi kepribadian adalah dengan menggunakan metode Pembelajaran Mesin. Beberapa penelitian telah mencoba melakukan deteksi kepribadian pengguna Twitter dengan berbagai metode seperti *Naïve Bayes, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), dan Convolutional Neural Network (CNN)*. Penelitian yang dilakukan dengan metode CNN memiliki kelebihan yang dapat meningkatkan hasil dalam skala besar [2]. Metode ini bekerja dengan baik untuk analisis gambar dan klasifikasi citra karena dapat mengekstrak area fitur dari informasi global, dan juga mampu mempertimbangkan hubungan di antara fitur-fitur ini. CNN mampu mengatasi masalah yang lebih kompleks dengan baik karena memiliki lebih banyak lapisan yang tersembunyi dibandingkan jaringan saraf lainnya [6]. Sementara itu, dalam analisis sentimen, CNN memiliki lapisan konvolusi untuk mengekstrak informasi dengan potongan teks yang lebih besar. CNN juga membutuhkan koneksi dan parameter yang lebih sedikit sehingga lebih mudah untuk dilatih. Selain itu, metode CNN yang digabungkan dengan fitur *word embedding* yang dihasilkan dari *Term Frequency-Inverse Document Frequency (TF-IDF)* dapat meningkatkan performa dalam proses klasifikasi [7]. Namun, performa machine learning dapat memiliki performa yang baik apabila data yang digunakan pada penelitian memiliki kualitas yang bagus. Pada penelitian ini, tentunya data kecenderungan kepribadian pada setiap pengguna berbeda-beda, sehingga kemungkinan adanya ketidakseimbangan pada proporsi kecenderungan kepribadian yang terdapat pada data cuitan pengguna bisa terjadi.

### Topik dan Batasannya

Terdapat beberapa batasan pada penelitian ini. Data yang digunakan yaitu cuitan Twitter yang diambil menggunakan Bahasa Indonesia. Batasan lain yaitu pengguna Twitter yang dipilih untuk dijadikan sebagai data pada penelitian ini memiliki lebih dari 1000 cuitan untuk setiap pengguna.

### Tujuan

Penelitian ini menggunakan metode *Convolutional Neural Network (CNN)* yang bertujuan untuk melakukan pendeteksian kepribadian pada pengguna media sosial Twitter. Selain itu tujuan dalam melakukan penelitian ini untuk mengetahui performa dari metode CNN dalam melakukan deteksi kepribadian pada media sosial Twitter.

### Organisasi Tulisan

Terdapat lima bagian dalam organisasi penulisan pada penelitian ini. Setelah pendahuluan, bagian dua menjelaskan beberapa studi yang terkait dengan penelitian ini serta metode yang digunakan. Selanjutnya pada bagian tiga yaitu penjelasan sistem yang dibangun pada penelitian ini. Setelah itu pada bagian empat menjelaskan tentang hasil pengujian dan analisis serta evaluasinya, dan bagian lima yang berisi kesimpulan dan saran dari penelitian ini.