

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Kecerdasan buatan atau biasa dikenal dengan Artificial Intelligence atau AI diartikan sebagai kecerdasan entitas ilmiah. Sistem seperti itu biasanya dianggap komputer. Kecerdasan diciptakan dan diintegrasikan ke dalam komputer sehingga bisa bekerja seperti manusia (Hoppe, Loetscher, Morey, & Bulling, 2018). Salah satu kecerdasan buatan yang dibuat manusia adalah *Natural Language Processing* atau bisa disebut NLP, yang bertujuan untuk melakukan proses pembuatan contoh komputasi berdasarkan bahasa, sebagai akibatnya bisa terjadi suatu hubungan antara manusia dengan komputer yang menggunakan perantara bahasa alami (Segalin et al., 2017). Model komputasi ini bisa bermanfaat buat keperluan ilmiah contohnya meneliti sifat-sifat berdasarkan suatu bentuk bahasa alami juga buat keperluan sehari-hari sehingga memudahkan komunikasi antara manusia dengan komputer. Salah satu *machine learning* yang dapat digunakan untuk mengenali kepribadian seseorang ialah *Natural Language Understanding* (NLU) yang juga merupakan salah satu sub bagian *Natural Language Processing* (NLP) (Adi, Tandio, Ong, & Suhartono, 2018).

Natural Language Understanding (NLU) atau bisa disebut sebagai pengertian Bahasa Alami merupakan sistem yang memprediksi kepribadian pengguna berdasarkan konten yang ditulis oleh pengguna di media sosial. NLU

dikembangkan oleh perusahaan besar *International Business Machines* (IBM) dalam berbagai bahasa, terutama bahasa Inggris. NLU memiliki banyak keunggulan di berbagai bidang bisnis. Misalnya, NLU membantu peneliti dan penyedia layanan mengembangkan produk dan fasilitas yang dipersonalisasi (Diana Ikasari, Suhendra, & Farida, 2018). Riset NLU telah banyak digunakan dan dikembangkan. Riset NLU terutama dilakukan dengan memeriksa pengguna media sosial seperti *Twitter*, Instagram, dan Facebook, yang menyebabkan perkembangan pesat pada media sosial dan menyediakan data pengguna dalam jumlah besar (Bahri, Saputra, & Wajhillah, 2017).

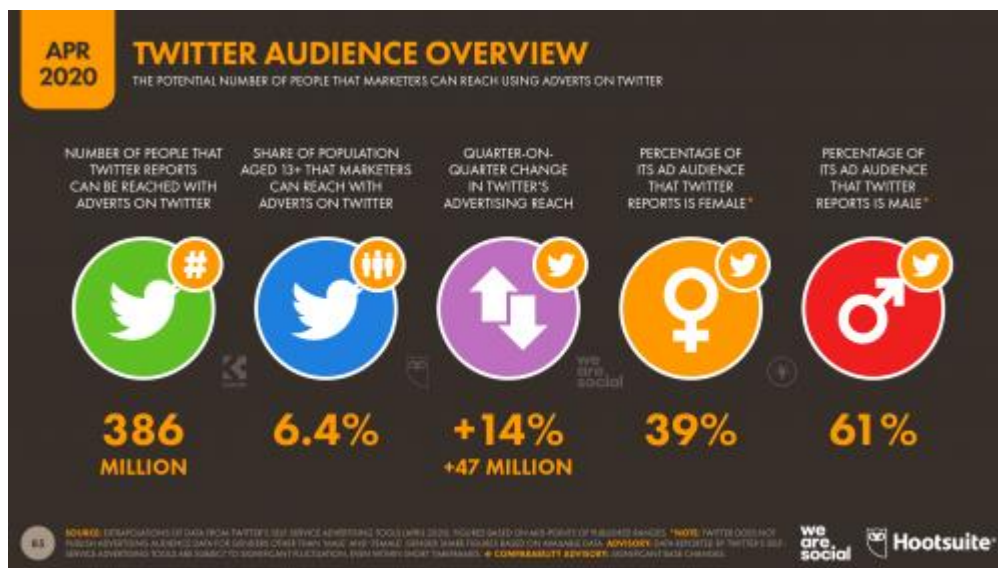


Gambar 1. 1 Statistik Pengguna Sosial Media *Twitter* Disetiap Negara

Sumber: (We Are Social & Hootsuite, 2020)

Gambar 1.1 adalah jumlah statistik pengguna media sosial *Twitter* disetiap negara dengan data terbarunya yaitu bulan April tahun 2020. Menurut data (We Are Social & Hootsuite, 2020), ada 11,8 juta pengguna sosial media *Twitter* di Indonesia dan naik 11% dari perhitungan Januari 2020 – April 2020. Angka ini menunjukkan

bahwa pengguna sosial media di Indonesia sangatlah banyak dan bisa dikatakan menjadi salah satu media sosial terpopuler di Indonesia. Dengan banyaknya pengguna sosial media *Twitter* dan posting konten massal yang dibuat oleh pengguna *Twitter* dari Indonesia menjadikan *Twitter* sebagai media sosial yang ideal untuk penelitian pengembangan *Natural Language Understanding*.



Gambar 1. 2 Jumlah Audience Overview Twitter

Sumber: (We Are Social & Hootsuite, 2020)

Gambar 1.2 merupakan jumlah *audience overview Twitter* dengan data terbarunya yaitu pada bulan April tahun 2020. Berdasarkan data (We Are Social & Hootsuite, 2020), *Twitter* telah mengalami pertumbuhan kuartalan terbesar dari *platform* ini. Data yang dirilis di alat periklanan layanan mandiri perusahaan menunjukkan bahwa, dibandingkan dengan Januari 2020, pengiklan kini dapat meningkatkan jumlah kunjungan di *Twitter* sebesar 47 juta, yang setara dengan peningkatan 14% dari bulan sebelumnya. Aktivitas pengguna *Twitter* terus meningkat dan saat ini menjadi situs jejaring sosial terpopuler di kalangan

pengguna Indonesia. Selain itu *Twitter* juga menyediakan API yang mudah digunakan dan merupakan media sosial yang paling banyak digunakan untuk melakukan penelitian NLU (Shelar & Huang, 2018). *Twitter* digunakan dalam basis harian oleh pengguna dari seluruh dunia untuk mengungkapkan opini dan berkomunikasi. Hal ini memberikan sebuah tantangan baru pada bidang klasifikasi sentimen dan opini. Tantangan ini berupa bagaimana cara mengklasifikasi sentimen dan opini dari teks yang pendek dan tidak terstruktur. *Twitter* yang merupakan salah satu situs media sosial terkenal dimana pengguna membagikan pesan terbaru sepanjang 140 karakter (*tweet*) tiap pesan. Penelitian ini juga bertujuan untuk mendapatkan emosi dari pengguna sosial media *twitter* berdasarkan *tweet* sepanjang 140 karakter yang dibagikan secara *online* serta dapat membantu perusahaan untuk mengambil keputusan saat merekrut karyawan baru.

Dari pengkajian yang telah dijabarkan, manfaat dalam penelitian *Natural Language Understanding* terhadap *Twitter* ini adalah untuk mengembangkan penelitian terhadap kepribadian pengguna menggunakan 6 sifat dasar manusia dengan cara mengambil *tweet* seseorang dan mengolah menjadi data sehingga bisa diprediksi kepribadian orang tersebut. Pada tahun 1970-an, psikolog Paul Eckman mengidentifikasi 6 sifat dasar manusia yang biasa dialami dalam kehidupan manusia. Sifat yang diidentifikasi adalah kebahagiaan (*joy*), kesedihan (*sadness*), jijik (*disgust*), ketakutan (*fear*), kejutan (*shock*), dan kemarahan (*angry*). Eckman kemudian memperluas daftar emosi dasarnya untuk dimasukkan seperti kesombongan (*arrogant*), rasa malu (*shame*), dan kegembiraan (*excitement*). (Capocasa, McCuen, Venier, & Rufo, 2016).

1.2. Rumusan Masalah

Berdasarkan latar belakang yang sudah dijelaskan, berikut ini rangkaian rumusan masalah yang akan dicari pemecahannya melalui penelitian ini.

1. Apakah NLU dapat mengklasifikasi emosi seseorang melewati sosial media *Twitter*?
2. Sifat dasar apa dengan hasil klasifikasi tertinggi dan terendah dari *five emotion* tersebut?

1.3. Batasan Masalah

Untuk menyelesaikan masalah di atas dan pembahasan dari penelitian ini tidak jauh dari masalah yang ada, maka batasan masalah sebagai berikut:

1. Penelitian berfokus pada layanan aplikasi *Twitter* dan data yang diambil hanya dari *Twitter*.
2. Data *tweet* yang diambil hanya berupa teks berbahasa Inggris.
3. Sifat dasar yang diklasifikasi hanya 5 dari 6 sifat dasar manusia yaitu kebahagiaan (*joy*), kesedihan (*sadness*), jijik (*disgust*), ketakutan (*fear*), dan kemarahan (*angry*) dan yang tidak ada adalah kejutan (*shock*).
4. *Tweet* yang diambil dari *Twitter* melewati *username* bukan dari *hashtag*.
5. Durasi *mining data Twitter* adalah 1000 *tweet* tiap pengguna.

6. Program klasifikasi emosi seseorang hanya bisa mengambil 1 *username* di setiap aplikasi dijalankan contohnya jika ingin mendeteksi 3 orang maka harus *run program* sebanyak 3x.

1.4. Tujuan dan Manfaat Penelitian

1.4.1. Tujuan Penelitian

1. Dapat memprediksi emosi seseorang melewati sosial media *Twitter* dengan Natural Language Understanding (NLU).
2. Memberikan hasil visualisasi penelitian dari metode Natural Language Understanding (NLU) pada analisis *Twitter*.

1.4.2. Manfaat Penelitian

1. Mendapatkan metode yang efektif dalam melakukan NLU yang mengkaji data dalam Bahasa Inggris.
2. Memperoleh data yang lebih baru dan luas untuk NLU sehingga bisa digunakan oleh peneliti lainnya di kemudian hari.