

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Kepribadian merupakan aspek psikologi dan faktor kunci yang mempengaruhi kinerja karyawan. Kepribadian telah dipelajari secara ekstensif dalam psikologi, dan ada beberapa teori tentang tes kepribadian. Tes kepribadian biasanya digunakan oleh perekrut sebagai parameter untuk menempatkan seseorang dalam sebuah bidang pekerjaan yang sesuai dengan kepribadian pekerja dan sesuai dengan kebutuhan perusahaan [1].

Media sosial adalah tempat untuk orang berbagi tentang diri mereka sendiri dan perspektif mereka tentang kehidupan. Salah satu media sosial yang paling populer di kalangan masyarakat Indonesia saat ini adalah Twitter. Terdapat 19,5 juta pengguna Twitter di Indonesia [2]. Twitter merupakan salah satu jenis media sosial *microblogging* di mana pengguna dapat menulis dan mempublikasikan aktivitas dan pendapat mereka. Beberapa penelitian telah dilakukan dengan menggunakan media sosial sebagai objek. Hal ini disebabkan pengakuan akan potensi banyak informasi (termasuk masalah kepribadian) yang dapat digali dari media sosial seperti Twitter [3].

*Big five personality* adalah salah satu teori pengujian kepribadian yang diakui secara internasional yang diajukan oleh psikolog terkemuka Louis Goldberg. Identifikasi kepribadian berdasarkan *big five personality traits* membutuhkan waktu yang lama dan tidak efisien karena diperlukan responden untuk mengisi kuesioner dengan total sekitar 20 sampai 360 pertanyaan [4]. Menurut sebuah penelitian kepribadian *big five* dipelajari dengan baik dan memiliki hubungan erat terhadap kehidupan dan kepuasan bekerja. Hal tersebut memberikan manfaat untuk perekrut menempatkan posisi pekerja yang tepat sesuai kepribadian *big five* [5].

Penelitian tentang identifikasi kepribadian berkembang dan menghasilkan metode baru seperti pengujian kepribadian otomatis dengan analisis data yang didapat dari media sosial. Pengukuran kepribadian otomatis ini memanfaatkan fitur linguistik dari sebuah teks di media sosial, hal ini juga dapat menjadi cara baru

untuk mengurangi biaya dan waktu yang diperlukan jika dibandingkan dengan pengukuran manual seperti wawancara dengan psikolog.

Saat ini tes penelitian mengenai tes kepribadian *big five* di media sosial sudah banyak dilakukan, hanya saja kebanyakan masih menggunakan *traditional machine learning*. Salah satunya adalah penelitian yang dilakukan oleh Rendy Putra Pratama. Penelitiannya dilakukan dengan tujuan membangun sebuah model prediksi kepribadian yang dimana dilakukan proses pengklasifikasian tweet untuk mengetahui sifat kepribadian berdasarkan kategori *big five* dengan menggunakan algoritma *random forest*. Penelitian tersebut menghasilkan akurasi 69,23% [6].

Penelitian mengenai perbandingan antara algoritma *Machine learning* dan algoritma *deep learning* telah banyak dilakukan, salah satunya pada penelitian yang dilakukan oleh Nuli Giarsyani. Pada penelitiannya tersebut membandingkan antara algoritma *machine learning* dan *deep learning* untuk mengklasifikasikan tweet berupa teks. Dan hasilnya adalah algoritma *deep learning* dengan akurasi terbesar adalah algoritma *Long Short-Term Memory (LSTM)* sebesar 99% dan algoritma *machine learning* dengan akurasi terbesar adalah *random forest* dengan nilai 98%. Penelitian tersebut membuktikan bahwa untuk pengklasifikasian teks metode *deep learning* dapat bekerja lebih baik dibandingkan dengan metode *machine learning* khususnya yaitu algoritma LSTM. Algoritma LSTM memiliki keunggulan karena dapat mengklasifikasikan informasi berdasarkan *data time series*. Menurut konsep ini, LSTM dapat mengingat dan menghapus data lama ketika tidak lagi diperlukan. Dengan demikian, manajemen informasi akan lebih lengkap sekaligus aktual [7].

Berdasarkan uraian latar belakang diatas, penelitian ini akan melanjutkan penelitian sebelumnya mengenai klasifikasi teks pada prediksi kepribadian *big five*. Penelitian kali ini akan menggunakan algoritma LSTM karena pada penelitian sebelumnya telah dibuktikan bahwa algoritma *deep learning* dapat menghasilkan akurasi lebih baik dibandingkan dengan algoritma *traditional machine learning*. Dengan permasalahan tersebut maka dirumuskan sebuah penelitian yang berjudul **“Prediksi Tipe Kepribadian *Big five* Pada Pengguna Twitter Dengan Algoritma *Long Short-Term Memory*”**. Penelitian ini diharapkan menghasilkan model pembelajaran mesin yang nantinya dapat dikembangkan lebih lanjut.

## 1.2. Rumusan Masalah

Rumusan masalah didasarkan pada latar belakang adalah sebagai berikut:

1. Bagaimana penerapan algoritma LSTM pada prediksi kepribadian *big five* pengguna Twitter?
2. Bagaimana kinerja algoritma LSTM pada prediksi kepribadian *big five* pengguna Twitter?

## 1.3. Tujuan Penelitian

Tujuan dari penelitian ini sebagai berikut:

1. Menganalisis dan mengimplementasikan algoritma LSTM untuk prediksi kepribadian *big five* pengguna Twitter.
2. Mengetahui kinerja algoritma LSTM pada prediksi kepribadian *big five* pengguna Twitter.

## 1.4. Manfaat Penelitian

Manfaat yang didapat dari penelitian ini sebagai berikut:

1. Dapat menambah wawasan dan ilmu peneliti serta dapat menerapkan hasil penelitian ini di kemudian hari.
2. Dapat menjadi bahan referensi bagi penelitian berikutnya mengenai topik *Text Classification* menggunakan *deep learning* menggunakan algoritma LSTM.

## 1.5. Batasan Masalah

Batasan masalah diperlukan agar penelitian ini dapat fokus terhadap hal-hal yang ingin dicapai dan agar ada Batasan terkait hal-hal tertentu yang sifatnya diluar batas. Berikut merupakan beberapa batasan masalah dari penelitian mengenai penerapan algoritma LSTM dalam prediksi kepribadian *big five* pengguna Twitter adalah:

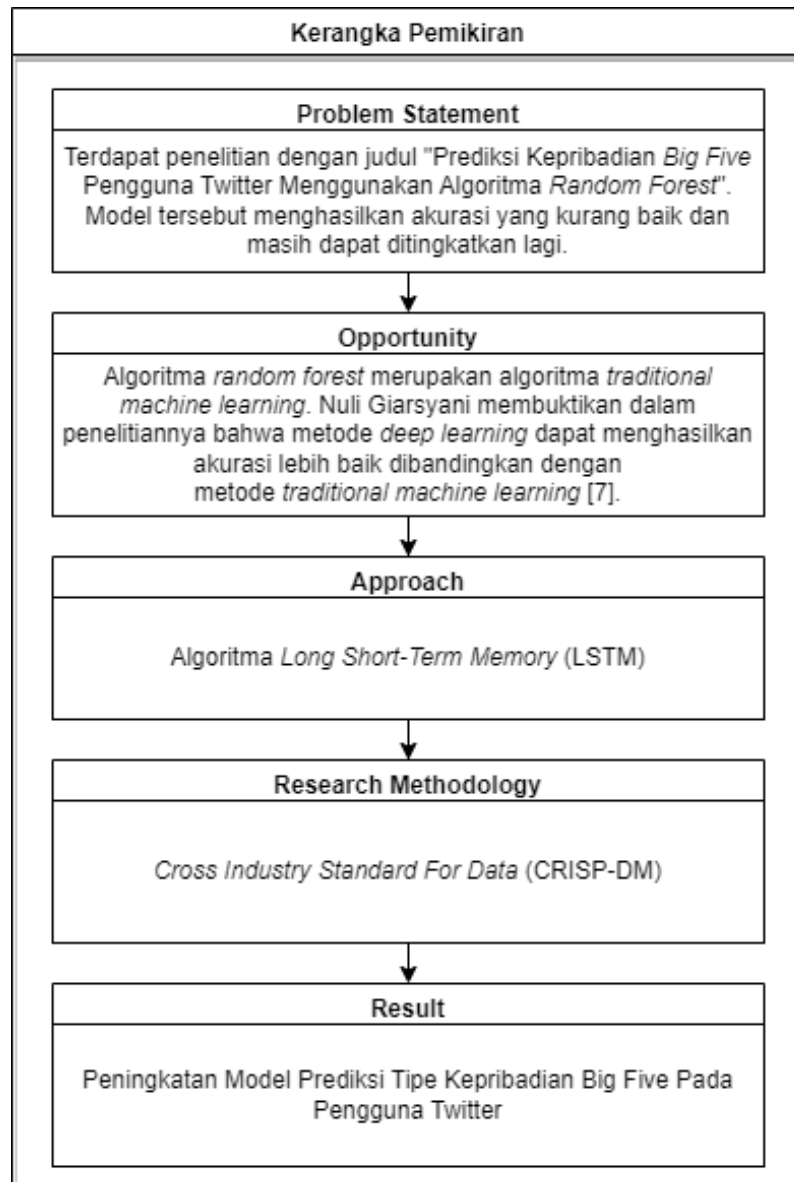
1. Penelitian ini melanjutkan penelitian sebelumnya yang dilakukan oleh Rendy Putra Pratama dan juga menggunakan *dataset* yang digunakan pada penelitian sebelumnya yang berisikan tweet yang telah dilabeli berdasarkan kategori kepribadian *big five*.
2. Prediksi kepribadian yang dilakukan hanya menggunakan data tweet.
3. Kepribadian diprediksi menggunakan teori *big five personality*.

4. Prediksi kepribadian yang dilakukan adalah kepribadian seseorang bersosial media.
5. Metode yang digunakan dalam penelitian ini menggunakan algoritma *Long Short-Term Memory*.
6. Tahapan CRISP-DM yang diaplikasikan pada penelitian ini tidak sampai pada tahap *deployment*.

#### **1.6. Kerangka Pemikiran**

Pernyataan masalah (*problem statement*) didapatkan dari penelitian yang dilakukan oleh Rendy Putra Pratama yang berjudul “Prediksi Kepribadian *Big Five* Pengguna Twitter Menggunakan Algoritma *Random Forest*”. Algoritma *random forest* yang digunakan pada penelitian tersebut merupakan algoritma *traditional machine learning*, hasil akurasi dari model yang dibuat pada penelitian tersebut masih dapat ditingkatkan lagi. Terdapat sebuah peluang/*opportunity* yaitu penelitian yang dilakukan Nuli Giarsyani yang membuktikan bahwa algoritma *deep learning* dapat menghasilkan akurasi yang lebih baik dibandingkan dengan algoritma *traditional machine learning* [7]. Untuk pendekatan yang dilakukan adalah dengan algoritma *Long Short-Term Memory* (LSTM) yang merupakan algoritma *deep learning* yang mampu memproses *sequential data* seperti data teks. Untuk metode penelitian (*research methodology*) yang digunakan adalah *Cross Industry Standard Process for Data Mining* (CRISP-DM) yang merupakan salah satu standard untuk *data mining* yang terpopuler pada tahun ini [8].

Hasil dari penelitian ini adalah prediksi tipe kepribadian *big five* pada pengguna twitter dengan algoritma *Long Short-Term Memory* (LSTM). Untuk kerangka pemikiran ini digambarkan pada Gambar 1.1 berikut:



Gambar 1.1 Kerangka Pemikiran

### 1.7. Sistematika penulisan

Sistematika penulisan pada penelitian ini dibagi ke dalam 5 bab. Pada setiap bab dijelaskan sesuai dengan tujuan dari pengembangan sistem tersendiri. Sistematika penulisan pada penelitian ini dapat dilihat sebagai berikut:

#### **BAB I : Pendahuluan**

Bab I menjelaskan tentang latar belakang penelitian dan dilanjutkan dengan rumusan masalah, tujuan, batasan, metode pengembangan dari sistem, kerangka sampai dengan kerangka pemikiran. Berikut pula metode penulisan yang disajikan.

## **BAB II: Kajian Literatur**

Bab II berisikan tentang pembahasan penelitian terdahulu serta konsep-konsep dan teori pendukung pada penelitian yang akan dilakukan.

## **BAB III: Metodologi Penelitian**

Bab III berisikan tentang mengenai metode yang digunakan dalam penyusunan tugas akhir. Metodologi penelitian disajikan berdasarkan analisis kebutuhan menggunakan metode CRISP-DM. Dalam metode tersebut beberapa tahapan yang terdapat pada Bab III ini adalah Pemahaman Bisnis, pemahaman Data, Persiapan Data, dan *Modeling Phase*.

## **BAB IV : Hasil dan Pembahasan**

Pada bab ini membahas mengenai hasil dari implementasi sistem itu sendiri seperti hasil dari perhitungan training dan pengujian yang dilakukan. Bab ini berisikan lanjutan tahapan CRISP-DM dari bab sebelumnya yaitu *Evaluation Phase*.

## **BAB V : Simpulan dan Saran**

Bab ini berisikan kesimpulan dari penelitian yang dilakukan serta saran yang direkomendasikan untuk peningkatan atau perbaikan dari penelitian ini.