

BAB 1. PENDAHULUAN

1.1. Latar Belakang

Pada saat ini, media sosial mengalami peningkatan penggunaan yang pesat dan telah menjadi bagian penting dari kehidupan sehari-hari. Media sosial merupakan sebuah tempat untuk pengguna dapat saling berinteraksi dan berbagi informasi [1]. Media sosial X merupakan salah satu platform yang banyak digunakan saat ini dan merupakan tempat untuk individu dapat mengekspresikan kepribadian diri mereka melalui cuitan secara bebas. Oleh karena itu kepribadian seseorang dapat diidentifikasi berdasarkan cuitan yang dibuat [2]. Kepribadian adalah pola pikir individu yang dipengaruhi oleh perilaku dan mencerminkan perbedaan dalam karakteristik setiap orang. Selain itu, kepribadian juga mencerminkan perilaku dan ciri khas unik dari masing-masing individu [3]. Pemahaman mengenai kepribadian melalui cuitan di media sosial X memiliki manfaat karena dapat membantu meningkatkan kinerja karyawan di lingkungan kerja [4]. Selain itu, analisis karakteristik kepribadian pengguna melalui teks di media sosial dapat membantu psikolog untuk mempercepat penilaian dan membantu perusahaan memilih kandidat yang sesuai [5].

Untuk mendeteksi kepribadian seseorang, terdapat beberapa model yang dapat digunakan seperti MBTI, DISC, dan *Big Five*. Namun, model *Big Five* adalah model yang paling populer dan akurat dalam memprediksi kepribadian seseorang, khususnya di kalangan peneliti [6] [7]. Model ini mengklasifikasikan kepribadian kedalam lima kategori yaitu *agreeableness*, *extroversion*, *conscientiousness*, *neuroticism*, dan *openness to experience* [8]. Penilaian *Big Five Personality* dalam mengklasifikasi kepribadian dapat dilakukan dengan menggunakan *Big Five Inventory* (BFI-44). BFI-44 adalah kuesioner yang terdiri dari 44 pertanyaan yang disusun dalam kalimat sederhana dengan rentang nilai dari 1 (sangat tidak setuju) hingga 5 (sangat setuju) [9].

Penelitian ini menggunakan metode IndoBERT untuk melakukan klasifikasi kepribadian berdasarkan cuitan di media sosial X. Menurut penelitian yang dilakukan oleh S. M. Isa [10], dalam mendeteksi berita palsu bahwa model IndoBERT

menghasilkan akurasi mencapai 94,66%, mengungguli metode TF-IDF+SVM (90%) dan TF-IDF+Naive Bayes (83%). Pada penelitian lain yang dilakukan oleh Gede Aditra Pradnyana [11], menunjukkan bahwa akurasi yang didapatkan dengan menggunakan metode IndoBERT-base mencapai 72% dengan F1-score sebesar 71%. Akurasi ini secara signifikan lebih tinggi dibandingkan dengan metode SVM yang memiliki akurasi yaitu 62% dan F1-score sebesar 57%. Selanjutnya pada penelitian yang dilakukan oleh Tri Ayu Syifa [12], menunjukkan bahwa metode IndoBERT lebih akurat dibandingkan dengan metode *logistic regression*, *adaboost* dan *decision tree*. Penelitian ini juga menunjukkan bahwa proses *stemming* dapat mempengaruhi kinerja dari model karena dapat mengubah makna dari suatu kata. Pada penelitian yang dilakukan oleh Lenggo Geni [13], metode IndoBERT memiliki akurasi yang cukup tinggi yaitu sekitar 83,5% dalam memprediksi analisis sentimen. Selanjutnya pada penelitian yang dilakukan oleh Benny Richardson [14], menunjukkan perbandingan metode *deep learning* antara IndoBERT dan Roberta. Hasilnya menunjukkan bahwa metode IndoBERT memiliki kinerja yang lebih baik dengan mendapatkan *f1-score* sebesar 89% dibandingkan dengan Roberta yang memiliki *f1-score* sebesar 87%.

1.2. Perumusan Masalah

Dengan mempertimbangkan latar belakang masalah yang ada, perumusan masalah pada penelitian ini adalah sebagai berikut:

1. Apa saja faktor yang dapat mempengaruhi kinerja dari model klasifikasi kepribadian menggunakan metode IndoBERT?
2. Bagaimana metode IndoBERT dapat mengklasifikasikan kepribadian berdasarkan cuitan pengguna di media sosial X?

1.3. Tujuan

Tujuan penelitian ini disusun berdasarkan rumusan masalah yang sudah diidentifikasi, yaitu sebagai berikut:

1. Untuk menganalisis dan mengevaluasi faktor yang dapat mempengaruhi kinerja dari model klasifikasi kepribadian menggunakan metode IndoBERT.

2. Untuk menganalisis metode IndoBERT dalam mengklasifikasikan kepribadian berdasarkan cuitan pengguna di media sosial X.

1.4. Batasan Masalah

Batasan yang didapatkan pada penelitian ini adalah:

1. Dataset merupakan teks yang berbahasa Indonesia.
2. Data diambil dari cuitan pengguna di media sosial X yang sudah mengisi dan menyetujui kuesioner BFI-44.
3. Penelitian berfokus untuk membangun model dengan menggunakan metode IndoBert.
4. Penelitian ini dilakukan dengan bahasa pemrograman python.

1.5. Hipotesis

Penelitian ini memiliki hipotesis sebagai berikut :

Premis 1: Metode IndoBERT-base dapat mengklasifikasikan *big five personality* dengan akurasi mencapai 72% dengan F1-score sebesar 71% [5].

Premis 2: Proses *stemming* dapat mempengaruhi kinerja dari model [12].

Hipotesis: Metode IndoBERT cocok digunakan untuk mengklasifikasikan kepribadian berdasarkan cuitan pengguna dengan melakukan proses *pre-processing* yang tepat.

1.6. Metode Penelitian

Rencana kegiatan yang pada penelitian ini dibagi menjadi 6 tahap sebagai berikut:

1. Studi Literatur

Tahap pertama adalah melakukan pengumpulan informasi terkait topik penelitian yang dilakukan. Pengumpulan informasi ini dilakukan dengan mengambil jurnal terkait yang ada di internet. Tujuan dari studi literatur adalah untuk mendapatkan referensi dari penelitian yang sudah dilakukan.

2. Pengumpulan Data

Pada tahap ini akan dikumpulkan dengan menyebarkan kuesioner BFI-44 dan meminta kesediaannya untuk akun X nya di analisis. Setelah itu, akan dilakukan proses *crawling data* dari X berdasarkan pengguna yang sudah mengisi kuesioner dan menyetujui akun X nya dianalisis.

3. Eksplorasi Data

Setelah data berhasil dikumpulkan, data ini akan dilakukan eksplorasi data. Eksplorasi data dilakukan dengan tujuan untuk mengetahui karakteristik dari data yang sudah didapatkan. Hasil dari eksplorasi data berupa visualisasi data.

4. Implementasi

Selanjutnya akan dilakukan implementasi dengan melakukan pembangunan model. Proses ini terdiri dari beberapa tahapan yaitu proses *pre-processing*, pembangunan model, dan evaluasi kinerja dari model yang sudah dibangun.

5. Analisis Hasil Penelitian

Pada tahap ini, proses analisis akan dilakukan untuk mengukur kinerja dari model yang sudah dibuat. Proses ini dilakukan dengan tujuan untuk menilai hasil dari penelitian yang dilakukan.