

BAB I PENDAHULUAN

1.1 Latar Belakang

Kepribadian ialah istilah latin ialah persona artinya kedok atau topeng. Kepribadian dapat memberi perbedaan karakter seseorang dalam berfikir, merasa dan berperilaku. Pada umumnya, cara menganalisis yang efektif untuk mengidentifikasi kepribadian seseorang salah satunya menggunakan model Big Five Personality agar dapat mengidentifikasi kepribadian seseorang dengan lebih informatif. Terdapat lima kepribadian didalamnya seperti Agreeableness, Openness, Extraversion, Conscientiousness, Neuroticism [13]. Banyak cara dalam menganalisis kepribadian seperti wawancara, observasi dan survey kuesioner, membuat analisis tersebut kurang dalam efektifitasnya. Untuk mengantisipasi terjadi kurangnya efektifitas dalam analisis terdapat studi yang menunjukkan bahwa tes kepribadian yang bisa diperoleh dari cara lain seperti menganalisis melalui media sosial seperti twitter yang dimana media tersebut merupakan platform untuk memposting kondisi, dan perasaan seseorang, dan platform tersebut sudah banyak pengguna didalamnya [5]. Suatu cara yang memanfaatkan sebagai analisis dataset dan melakukan prediksi pada pola terdapat di suatu data adalah data Mining. Didalam data mining mampu diperoleh melalui beberapa cara salah satunya adalah Classification. Sedangkan teknik pada pengumpulan datadatanya adalah klasifikasi, beberapa mekanisme yang digunakan pada data mining salah satunya seperti Random Forest. Adapun beberapa penelitian yang mengangkat mengenai salah satu model klasifikasi yang digunakan dalam penelitian mereka, sebagaimana penelitian yang dilakukan dengan memakai Random Forest serta Multivariate Adaptive Regression Spline (MARS) binary response merupakan hasil, dimana variabel yang mempunyai pengaruh terbaik pada status HIV/AIDS di Surabaya yaitu umur, setelah itu profesi, pernah ditahan akibat kejadian NAPZA, status nikah dan pemakaian jarum tertentu. Dimana hasil akurasi yang diperoleh sebesar 80,28%, lalu RF MARS dengan 91,00% serta hasil akurasi tertinggi yang diperoleh metode Random Forest yaitu dengan akurasi 97,80% [15]. Lalu dalam studi ini penulis akan berupaya melakukan klasifikasi menggunakan metode Random Forest dengan menerapkan Linguistic Inquiry and Word Count (LIWC) untuk mengetahui kepribadian seseorang dalam media sosial twitter . Terdapat 2 skenario di dalamnya, skenario pertama dengan melakukan pengumpulan setelah itu dilakukan analisis.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, maka permasalahan yang muncul dapat dirumuskan:

1. Bagaimana cara untuk mendeteksi personality seseorang menggunakan metode Random Forest ?
2. Bagaimana cara mengukur Performansi sistem yang dapat mendeteksi personality?
3. Bagaimana faktor-faktor yang berpengaruh dalam deteksi personality seseorang?

1.3 Tujuan

Berdasarkan latar belakang diatas, maka tujuan dari penelitian ini adalah sebagai berikut :

1. Mendeteksi personality melalui aktivitas twitter menggunakan metode Random Forest
2. Merancang sistem yang dapat mendeteksi personality melalui aktivitas di twitter memakai metode Random Forest.
3. Mengetahui factor apa saja yang mempengaruhi personality seseorang berdasarkan Big Five Model.

1.4 Batasan Masalah

Batasan masalah yang dibahas dalam penelitian ini adalah sebagai berikut :

1. Pengumpulan Kosa Kata yang dipakai adalah bahasa Indonesia dan Inggris.

2. Klasifikasi data memakai Random Forest.
3. Media sosial yang dipakai ialah Twitter.
4. Objek penelitian yang diteliti ialah Prediksi Personal yang dilakukan di twitter.