

## 1. Pendahuluan

Kepribadian adalah salah satu syarat mutlak bagi manusia untuk memancarkan eksistensinya di dunia, baik dalam dirinya sendiri maupun dalam kehidupan sosialnya. Kepribadian membuat seseorang menjadi unik dan berbeda dari satu orang ke orang lainnya, atau biasa disebut *individual differences*[1]. Dikarenakan kepribadian merupakan hal yang unik yang dimiliki setiap manusia, tidak jarang kepribadian juga dijadikan sebagai salah satu syarat dalam berbagai perekrutan seperti pendaftaran sekolah dan pendaftaran kerja melalui tes kepribadian. Hal ini dikarenakan adanya hubungan antara kepribadian dengan kinerja karyawan, seperti yang dijelaskan oleh Vianny Jeniston Delima pada penelitiannya tahun 2020 tentang dampak kepribadian terhadap prestasi kerja karyawan pada Batticaloa Teaching Hospital bahwa kepribadian berkorelasi positif dengan prestasi kerja[2]. Salah satu cara untuk mengetahui kepribadian calon karyawan pada perusahaan adalah dengan menggunakan jasa HRD atau *Human Resource Department*, namun penggunaan sistem manual seperti wawancara dan mengisi kuisioner terkait kepribadian menjadi tidak efektif dari segi waktu dan biaya. Selain itu, ada juga masalah yang muncul ketika tes dilakukan menggunakan sistem menjawab pertanyaan yang disediakan di kertas atau media lain menjadi tidak akurat karena kurangnya kemampuan calon karyawan dalam mengenali dirinya sendiri. Oleh karena itulah, dibutuhkan sistem yang dapat digunakan untuk mendeteksi kepribadian agar lebih hemat dari segi waktu dan biaya.

Metode Big Five Personality merupakan salah satu metode untuk menentukan kepribadian seseorang yang terbagi menjadi 5 dimensi yaitu *ekstraversi*, *agreeableness*, *conscientiousness*, *neuroticism*, dan *openness to experience* seperti yang disampaikan oleh Raad dan Mlacic dalam bukunya yang berjudul *The Big Five Personality Factors : The psycholexical approach to personality* tahun 2020 [3]. Penelitian juga telah dilakukan oleh Nasroh pada tahun 2017 untuk mengetahui pengaruh kepribadian dengan menggunakan metode *big five* terhadap kinerja karyawan. Dari penelitian tersebut menghasilkan kesimpulan bahwa terdapat dua dimensi yang berhubungan signifikan dengan kinerja karyawan yaitu *agreeableness* dan *neuroticism*, sedangkan tiga dimensi lainnya memiliki hubungan yang tidak signifikan[4].

Seerti yang sudah dijelaskan sebelumnya bahwa terdapat beberapa kekurangan jika menggunakan sistem kuisioner, sehingga dibutuhkan cara lain yang dapat digunakan untuk mendeteksi kepribadian seorang calon karyawan seperti yang telah dilakukan oleh Rahma Indira pada tahun 2021 dengan menggunakan metode *Long Short-Term Memory (LSTM)* dan fitur *Word2Vec*[5]. Pada penelitian tersebut menunjukkan bahwa deteksi kepribadian dapat dilihat dari cuitan twitter dan membutuhkan beberapa langkah hingga selesai antara lain pengumpulan data, *preprocessing data*, ekstraksi fitur, klasifikasi dan yang terakhir evaluasi. Dari hasil penelitian tersebut diketahui bahwa model dengan 10 tweet gabungan menghasilkan tingkat akurasi yang lebih tinggi dan penelitian ini membutuhkan 2 data dalam jumlah yang lebih besar, karena disebutkan bahwa kecilnya tingkat akurasi dan adanya *overfitting* karena datanya sedikit. Pada penelitian lainnya untuk mendeteksi kepribadian *Big Five Personality* dapat menggunakan metode C4.5 seperti yang telah dilakukan oleh Shantika Valerin Therik tahun 2021 pada jurnalnya yang berjudul *Deteksi Kepribadian Big Five Pengguna Twitter Dengan Metode C4.5*[6]. Pada penelitian tersebut didapatkan *baseline* dari akurasi perilaku sosial memiliki tingkat akurasi sebesar 44.82%. Kemudian dibandingkan dengan penambahan data TF-IDF dan LIWC dengan SMOTE yang keduanya dilakukan dengan menerapkan teknik *hyperparameter tuning*. Penambahan TF-IDF dan LIWC memiliki hasil akurasi sebesar 62.06%, mengalami kenaikan sebesar 17.24% dari *baseline*. Sedangkan dengan penambahan data TF-IDF, LIWC, dan SMOTE mendapatkan hasil akurasi sebesar 76.92%, dengan kenaikan 32.1% dari *baseline*.

Penelitian di bidang NLP atau *Natural Language Processing* yang menggunakan bahasa indonesia tidak tersedia cukup banyak, karena kurangnya sumber data yang ada padahal bahasa indonesia merupakan bahasa yang digunakan sekitar 199 jutaan orang dengan peringkat ke 11 bahasa yang sering digunakan didunia tahun 2022 [7]. Oleh karena itu diciptakan model pretrained yang bernama *IndoBERT* [8]. *IndoBERT* merupakan salah satu variasi BERT (*Bidirectional Encoder Representations from Transformers*) dengan mengikuti BERT-Base (*uncased*) yang memiliki 12 tugas dengan train lebih dari 220M kata yang memiliki main resource pada Indonesia Wikipedia (74M kata), artikel Kompas, Tempo dan Liputan6 (total 55M) dan Indonesia Web Corpus (90 M kata). *IndoBERT* sendiri baru diterbitkan papernya pada September 2020 lalu dengan kumpulan data dari *Indo4B* yang digunakan untuk pelatihan monolingual kontekstual pra terlatih yang tersedia untuk umum seperti teks, blog, berita dan situs web [8]. Penggunaan metode *IndoBERT* pada penelitian deteksi penggunaan kalimat abusif pada teks berbahasa indonesia yang dilakukan oleh Hadiyan Putra pada tahun 2021 juga berhasil mendapatkan tingkat akurasi tertinggi yang dibandingkan dengan penggunaan metode KNN, SVM dan Naive Bayes [9]. Pada percobaan pertama, KNN dan SVM diketahui hanya dapat mengklasifikasikan data terhadap kelas mayoritas dan Naive Bayes sudah mampu mengklasifikasikan kelas minoritas namun memiliki akurasi yang lebih rendah dari pada *IndoBERT* yakni 0.3391 pada kelas 1, 0.0863 pada kelas 2 dan 0.2368 pada kelas ketiga.

Penelitian ini akan menggunakan metode indoBERT untuk mendeteksi kepribadian berdasarkan Big Five Personality dengan mempertimbangkan fitur berupa cuitan dan data label dari responden yang akan menghasilkan jenis dimensi dari kepribadian responden tersebut. Pemilihan penggunaan twitter sebagai sosial media yang dipilih untuk mendapatkan dataset karena twitter merupakan sosial media yang memiliki pengguna terbanyak ke dua di dunia setelah facebook dan Indonesia menjadi negara kelima dengan user twitter terbanyak di dunia setelah Brazil [10]. Hasil dari penelitian diharapkan dapat mengetahui cara mendeteksi personality dari tweet pengguna dengan big five personality dan metode IndoBERT.