

1. Pendahuluan

Kepribadian Big five merupakan salah satu metode yang dikenal dalam dunia psikologi untuk menginterpretasikan kepribadian seseorang, terutama untuk menemukan hubungan kepribadian dengan lingkungan pekerjaan. Kepribadian *big five* terdiri dari *Openness to Experience (O)*, *Conscientiousness (C)*, *Extraversion (E)*, *Agreeableness (A)*, dan *Neuroticism (N)* (Costa P. Dkk, 1991), kepribadian O memiliki imajinasi yang aktif, kepekaan terhadap estetika, kepedulian terhadap perasaan pribadi, ketertarikan terhadap perbedaan, keingintahuan intelektual, dan kebebasan berpendapat, kepribadian C berhubungan erat dengan mengendalikan *impulse*, pengendalian diri demi perencanaan yang matang, pengaturan, dan pengerjaan tugas tugas. Kepribadian E percaya diri, aktif, cerewet, optimis, serta menyukai kesenangan dan selalu merasakan ceria secara alami. Kepribadian A mengutamakan orang lain, simpatik terhadap orang lain, dan suka menolong. Kepribadian N cenderung mengalami perasaan negatif seperti ketakutan, kesedihan, rasa canggung, kemarahan, serta rasa bersalah dan suka benci.

Kepribadian berhubungan dan mempengaruhi beberapa aspek dari linguistik. Prediksi berdasarkan linguistik dilakukan dengan menganalisis pemilihan kata kata dan letak kata tersebut di dalam katagori yang di tentukan sesuai dengan bahasa yang digunakan. Analisis linguistik telah dilakukan terhadap beberapa Profil media sosial dan penggunaan bahasa sehari hari, pesan singkat. Para psikologi dengan menemukan korelasi berbagai variabel linguistik dengan kepribadian. Beberapa perusahaan khususnya perusahaan industri menengah ke atas telah menggunakan media sosial untuk mempertimbangkan penerimaan pegawai baru, selain hasil test psikologi formal yang selalu dilakukan (*CareerBuilder*, 2012). Berdasarkan hal ini, kepribadian seseorang dapat diprediksi berdasarkan informasi seseorang melewati akun media sosial, seperti *facebook* atau *Twitter*. Pada penelitian ini bertujuan membangun model prediksi kepribadian *Big five personality* dari penggunaan twitter dengan menggunakan *k-Nearest Neighbor (kNN)* pada penelitian [1] telah dilakukan penelitian tentang prediksi kepribadian dari *Twitter*, menggunakan metode regresi. Untuk kasus prediksi kepribadian pada peneliti [2] klasifikasi kepribadian berdasarkan teks twitter menggunakan *Naive Bayes*, kNN dan SVM.

Tujuan penelitian ini yaitu untuk mengetahui kepribadian pengguna twitter yang telah di-*survey* dan dilebelkan, maka seberapa besar akurasi yang dihasilkan oleh metode kNN dan TF-IDF yang diaplikasikan pada analisis *Big five Personality*.