

1. Pendahuluan

Pada tahun 2020, dunia dihadapkan dengan ancaman virus corona yang mengubah cara dan berkurangnya interaksi manusia secara nyata yang drastis. Dampaknya terasa luas, termasuk meningkatnya jumlah pengguna aktif media sosial yang mencerminkan keberagaman manusia [12]. Di antara beragam platform media sosial, Twitter menjadi salah satu yang paling populer, tempat di mana pengguna berbagi cuitan informasi termasuk pesan yang mencerminkan kepribadian individu[1]. Kepribadian menjadi aspek penting dalam studi manusia, dan Deteksi Kepribadian Big Five menjadi fokus penelitian yang menarik. Dimensi kepribadian Big Five mencakup kestabilan emosional, keakraban, keterbukaan, kesadaran, dan ekstroverasi. Cuitan pengguna twitter memberikan kesempatan untuk melakukan analisis dan untuk mengevaluasi untuk memahami sentimen atau emosi yang terkandung dalam teks, termasuk pesan di media sosial Twitter.

Dalam upaya untuk mencapai tujuan penelitian ini memilih menggunakan algoritma machine learning, khususnya algoritma C4.5 dan Adaboost. Kedua algoritma ini dipilih karena kinerja dan keefektifannya telah terbukti dalam beberapa penelitian sebelumnya. Algoritma C4.5 dipilih karena mudah diinterpretasi, dapat mengatasi data kategorikal seperti teks tweet pada Twitter, dan berkontribusi dalam analisis sentimen dan deteksi kepribadian Big Five. Di sisi lain, Adaboost adalah metode ensemble learning yang meningkatkan performa model lemah menjadi model yang lebih kuat dengan menggabungkan beberapa model lemah. Kelebihan Adaboost meliputi kemampuan mengatasi data tidak seimbang (yang sering terjadi dalam data media sosial), efisien, dan mudah diimplementasikan. Kedua algoritma ini diharapkan dapat memberikan hasil analisis sentimen yang lebih akurat dalam mendeteksi kepribadian Big Five pengguna Twitter.

Penelitian ini menggunakan pendekatan cerdas dengan menggabungkan proses analisis sentimen dengan algoritma C4.5 dan Adaboost. Dengan demikian, diharapkan mampu menghasilkan model prediktif berdasarkan data tweet yang telah dianotasi dengan dimensi kepribadian. Pendekatan ini memberikan wawasan yang lebih mendalam tentang preferensi, sikap, dan perilaku pengguna Twitter. Analisis sentimen yang ditingkatkan juga akan meningkatkan akurasi dan validitas deteksi kepribadian Big Five. Selain itu, penelitian ini juga akan menyelidiki dan membandingkan dua algoritma klasifikasi utama, yaitu algoritma C4.5 dan Adaboost, dalam hal interpretabilitas dan performa. Algoritma C4.5 merupakan algoritma pohon keputusan yang mudah dipahami oleh manusia karena menghasilkan struktur pohon keputusan yang jelas dan sederhana. Di sisi lain, Adaboost merupakan teknik ensemble learning yang bertujuan meningkatkan kinerja model klasifikasi dengan menggabungkan beberapa model lemah menjadi model yang lebih kuat. Eksperimen dan evaluasi kinerja dari kedua algoritma ini akan dilakukan untuk memberikan gambaran yang komprehensif tentang efektivitas dan keunggulan masing-masing metode.

Latar Belakang

Keberagaman manusia saat ini menjadi topik penelitian yang relevan dan menarik karena setiap individu memiliki atribut kepribadian yang unik, yang dapat mempengaruhi perilaku dan cara berkomunikasi. Salah satu platform yang menonjol dalam menunjukkan keberagaman manusia adalah media sosial, khususnya Twitter. Media sosial ini telah menyediakan banyak informasi tentang bagaimana orang berkomunikasi menggunakan ucapan atau mesin tik. Oleh karena itu, memahami bagaimana media sosial, terutama Twitter, dapat digunakan untuk menganalisis kepribadian seseorang menjadi hal penting. Dalam konteks ini, Teori kepribadian Big Five memberikan kerangka

kerja yang kuat untuk memahami kepribadian, dengan mencakup lima kualitas utama: *Extraversion*, *Agreeableness*, *Openness*, *Conscientiousness*, dan *Neuroticism*. Melalui analisis media sosial, khususnya Twitter, kita dapat mengidentifikasi pola-pola bahasa yang terkait dengan kepribadian Big Five pengguna. Memanfaatkan algoritma machine learning, seperti algoritma C4.5 dan Adaboost untuk mengklasifikasikan kepribadian Big Five berdasarkan *tweet* pengguna Twitter. Pendekatan ini merupakan pengembangan dari penelitian yang ada sebelumnya. Algoritma C4.5 dan Adaboost dipilih karena keduanya adalah algoritma ensemble yang dapat dikombinasikan dengan metode Bagging untuk meningkatkan performa klasifikasi. Penelitian sebelumnya juga telah menunjukkan bahwa penggabungan algoritma C4.5 dengan metode Bagging dan Adaboost meningkatkan akurasi prediksi.

Analisis sentimen sendiri memiliki definisi sebagai proses penggunaan algoritma dan teknik komputasional untuk mendeteksi, mengidentifikasi, dan memahami sentimen atau emosi yang terkandung dalam teks atau data. Tujuannya adalah untuk mengklasifikasikan teks menjadi kategori sentimen yang berbeda, seperti positif, negatif, atau netral, serta mungkin juga kategori yang lebih rinci seperti senang, sedih, marah, dan lain-lain. Adapun hubungan Kepribadian pada Analisis Sentimen adalah dalam analisis sentimen di bidang pembelajaran mesin, kepribadian juga dapat memainkan peran penting dalam menginterpretasikan hasil. Misalnya, seseorang dengan kepribadian *skeptis* mungkin cenderung meragukan akurasi atau validitas algoritma analisis sentimen, sementara seseorang dengan kepribadian *optimis* mungkin lebih mudah menerima hasil tersebut. Hal ini dilakukan dengan menganalisis pola-pola bahasa yang digunakan oleh pengguna Twitter dalam *tweet* mereka. Hasil analisis sentimen kemudian digunakan untuk menentukan tingkat Kepribadian seseorang dengan model Big Five. dari pengguna Twitter.

Memberikan gambaran yang lebih akurat tentang kepribadian pengguna Twitter berdasarkan model Big Five dengan menganalisis sentimen dari *tweet* pengguna. Hasil analisis sentimen ini diharapkan dapat memberikan kontribusi bagi peneliti lain yang membutuhkan pemahaman pandangan orang tentang topik atau peristiwa tertentu. Metodologi penelitian ini melibatkan pengambilan sampel *tweet* dari pengguna Twitter yang telah diidentifikasi kepribadiannya menggunakan Big Five. Data *tweet* kemudian akan diproses menggunakan algoritma C4.5 dan Adaboost untuk menganalisis sentimen dan mendeteksi pola-pola bahasa yang terkait dengan kepribadian Big Five.

Topik dan Batasannya

Dalam penelitian ini, deteksi kepribadian Big Five hanya diterapkan pada *tweet* yang berisi teks dan tidak mempertimbangkan media lain seperti gambar atau video. Fokus analisis hanya terbatas pada konten teks yang terdapat dalam *tweet* pengguna Twitter. Selain itu, penelitian ini bertujuan untuk membandingkan dan menganalisis performa kedua algoritma tersebut dalam mendeteksi sifat pada kepribadian Big Five. Selanjutnya, model yang dikembangkan hanya diterapkan pada *tweet* yang berbahasa Indonesia dan tidak dapat digunakan untuk *tweet* dalam bahasa lain. Penelitian ini memiliki fokus untuk mengevaluasi dan menganalisis performa algoritma C4.5 dan Adaboost dalam mendeteksi sifat OCEAN kepribadian Big Five pada pengguna Twitter. Performa dari kedua algoritma ini diukur menggunakan metrik evaluasi yang sesuai, seperti akurasi, presisi, recall, dan F1-score. Selanjutnya menentukan algoritma yang memberikan akurasi terbaik dalam memprediksi kepribadian Big Five dengan membandingkan kinerja kedua algoritma ini dapat diketahui algoritma mana yang lebih efektif dan akurat dalam

mengidentifikasi *Openness, Conscientiousness, Extraversion, Agreeableness, dan Neuroticism* dari konten *tweet* pengguna.

Tujuan

Tujuan dari penelitian ini adalah untuk mengembangkan dan menerapkan metode analisis sentimen menggunakan algoritma C4.5 dan Adaboost pada *tweet* yang di bagikan pengguna Twitter. Penelitian ini bertujuan untuk mengidentifikasi pola sentimen dari postingan dan aktivitas pengguna Twitter untuk menganalisis hubungan antara sentimen yang diekspresikan oleh pengguna dengan kepribadian *Openness, Conscientiousness, Extraversion, Agreeableness, dan Neuroticism*.

Penelitian ini akan melibatkan pengumpulan data dari pengguna Twitter yang relevan dengan topik penelitian, melakukan pelabelan data untuk analisis sentimen, dan melalui tahap preprocessing, normalisasi, serta seleksi fitur untuk mempersiapkan data sebelum dipecah menjadi data uji dan data latih. Dua algoritma yang akan digunakan dalam penelitian ini adalah algoritma C4.5 dan Adaboost. Penggunaan kedua algoritma ini akan dibandingkan dalam pengenalan sentimen dan pengenalan ciri-ciri kepribadian lima besar dari data Twitter. Hasil analisis sentimen dan pengenalan kepribadian lima besar akan dievaluasi untuk menilai kinerja dan akurasi keduanya. Dalam akhir penelitian, diharapkan dapat diperoleh kesimpulan mengenai efektivitas algoritma C4.5 dan Adaboost dalam menganalisis sentimen dan mengenali kepribadian Big five pada data pengguna Twitter. Hasil penelitian ini diharapkan dapat memberikan wawasan baru tentang penggunaan media sosial sebagai sumber data untuk mengidentifikasi ciri-ciri kepribadian dan pola perilaku pengguna di platform Twitter.

Tabel 1 Keterkaitan antara tujuan, pengujian dan keterangan

Tujuan	Pengujian	Keterangan
- Memberikan gambaran yang lebih akurat tentang kepribadian pengguna Twitter berdasarkan model Big Five dengan menganalisis sentimen dari <i>tweet</i> pengguna. Bertujuan untuk memahami dan mengklasifikasikan kepribadian berdasarkan lima kualitas utama yang mencakup <i>Extraversion, Agreeableness, Openness, Conscientiousness, dan Neuroticism</i> .	Pengujian dilakukan dengan mengambil sampel <i>tweet</i> dari pengguna Twitter yang telah diidentifikasi kepribadiannya menggunakan tes Big Five. Data <i>tweet</i> kemudian diproses menggunakan algoritma C4.5 dan Adaboost untuk menganalisis sentimen dan mendeteksi pola bahasa yang terkait dengan kepribadian Big Five. Pengujian ini akan membantu memastikan apakah pendekatan dan algoritma yang digunakan dapat memberikan hasil yang akurat dan relevan dengan tujuan penelitian.	Mencakup detail tentang metodologi yang digunakan, termasuk bagaimana data <i>tweet</i> diambil, bagaimana algoritma C4.5 dan Adaboost diterapkan, serta bagaimana hasil analisis sentimen dianalisis untuk mengidentifikasi pola-pola bahasa terkait kepribadian Big Five. Keterangan juga akan mencakup interpretasi hasil dan kesimpulan yang diambil dari penelitian ini, serta implikasi dari temuan tersebut.