

## **Abstrak**

Rasa ingin tahu untuk memprediksi kepribadian seseorang melalui sosial media twitter bukanlah hal yang baru lagi. Setiap *tweet* atau kicauan yang ditulis dapat memberikan informasi mengenai kepribadian seseorang. Permasalahannya adalah bagaimana cara mengklasifikasi teks yang ada di sosial media Twitter ke dalam kelas-kelas yang akan dibuat dengan nilai performansi yang baik. Pada penelitian tugas akhir ini, penulis membangun sistem untuk mengklasifikasi kepribadian seseorang di sosial media twitter menggunakan metode klasifikasi decision tree C4.5 dengan metode pembobotan TF-RF dan TF-CHI<sup>2</sup>. Hal lain yang membedakan dengan penelitian ini adalah pembobotan setiap kata menggunakan metode pembobotan *TF-RF* dan *TF-CHI*<sup>2</sup> dengan penambahan fitur baru untuk pendekatan berdasarkan perilaku sosial pengguna seperti jumlah karakter pada *tweet*, rata-rata karakter pada setiap *tweet*, rata-rata kata pada *tweet*, media URL yang membaca seberapa banyak pengguna mengunggah foto atau video, tanda baca yaitu menghitung jumlah tanda tanya (?) dan tanda seru (!) pada setiap *tweet*, menghitung huruf besar, dan emoji pada *tweet* yang dapat dideteksi sebanyak 2.552 karakter yang berbeda-beda, dan yang terakhir adalah mencoba memadukan pendekatan kepribadian berdasarkan perilaku sosial dengan pendekatan linguistik. Dari hasil percobaan pada rasio data latih 90% dan data uji 10% (90:10) didapatkan hasil akurasi 65.72% melalui perpaduan fitur kepribadian berdasarkan perilaku sosial dengan pendekatan linguistik dengan pembobotan TF-RF.

**Kata kunci : Decision Tree C4.5, Term Weighting, TF-RF, TF-CHI<sup>2</sup>, Big Five**