

Klasifikasi Kepribadian Berdasarkan *Big Five Personality* Menggunakan Metode *Fuzzy Decision Tree* Dengan Algoritma C4.5

Siti Aisah^{1)*}, Fajri Rakhmat Umbara²⁾, Herdi Ashaury³⁾

¹⁾²⁾³⁾ Jurusan Informatika, Fakultas Sains dan Informatika, Universitas Jenderal Achmad Yani

Correspondence author : Siti Aisah, sitiaisyah18@if.unjani.ac.id, Cimahi, Indonesia

DOI: <https://doi.org/10.37012/jtik.v8i1.1110>

Abstrak

Kepribadian merupakan sifat alami seseorang mencakup tingkah laku, cara berbicara, bertindak dalam menghadapi sesuatu. Pada setiap orang memiliki kepribadian yang berbeda-beda dalam soal kualitas dan kuantitas. Banyak penelitian sudah dilakukan untuk klasifikasi kepribadian yang menggunakan media sosial sebagai sumber informasi. Memahami kepribadian seseorang dapat melalui tweet pada pengguna twitter dengan hasil klasifikasi gabungan antara perilaku sosial dan linguistik. Nilai akurasi penilaian kepribadian dapat ditingkatkan jika menggunakan salah satu dari pembobotan yang mempengaruhi kata-kata dari suatu tweet pengguna. Namun data yang digunakan pada penelitian ini memiliki atribut tipe diskrit dan kontinu sehingga membutuhkan metode yang dapat menangani permasalahan pada klasifikasi kepribadian. Pada penelitian ini klasifikasi kepribadian berdasarkan *Big Five Personality* sehingga menemukan kepribadian baru seperti *Extraverted*, *Serious*, *Dependable*, *Lively* dan *Responsible*. Data yang digunakan akan diekstraksi menggunakan metode *Fuzzy Decision Tree* dengan Algoritma C4.5. Data kepribadian yang digunakan terdiri dari 8 variabel. Model *Fuzzy Decision Tree* diimplementasikan menggunakan *pruning* atau pemangkasan, sehingga ada kemungkinan aturan yang dihasilkan memiliki akurasi yang tinggi.

Kata kunci : Kepribadian, *Big Five Personality*, *Fuzzy Decision Tree*, Algoritma C4.5

Abstract

Personality is the nature of a person including behavior, way of speaking, acting in the face of something. Everyone has a different personality in terms of quality and quantity. Many studies have been conducted for personality classification using social media as a source of information. Understanding a person's personality can be through tweets on Twitter users with the results of a combined classification of social and linguistic behavior. The accuracy value of personality assessment can be increased if one of the weights that affect the words of a user's tweet is used. However, the data used in this study has discrete and continuous type attributes, so it requires a method that can deal with problems in personality classification. In this study, personality classification is based on the Big Five Personality so as to find new personalities such as Extraverted, Serious, Dependable, Lively and Responsible. The data used will be extracted using the Fuzzy Decision Tree method with the C4.5 Algorithm. The personality data used consists of 8 variables. The Fuzzy Decision Tree model is implemented using pruning, so there is a possibility that the resulting rules have high accuracy.

Keywords: Personality, *Big Five Personality*, *Fuzzy Decision Tree*, Algorithm C4.5

PENDAHULUAN

Kepribadian merupakan gambaran unik dari seseorang (Yusup et al., 2021). Setiap individu memiliki kepribadian yang berbeda-beda dalam soal kualitas dan kuantitas. Dalam bukunya, Adolf Heuken S.J (1989 : 10) menegaskan “ Kepribadian adalah pola menyeluruh semua kemampuan, perbuatan serta kebiasaan-kebiasaan seseorang baik jasmani, mental,

rohani, emosional maupun yang sosial”. Kepribadian orang tersebut dapat menyatakan potensi yang dimiliki, dan kelompok yang sesuai dengan orang tersebut. Mengetahui kepribadian seseorang dapat digunakan untuk bahan penilaian personality seseorang dalam kehidupan sehari-hari.

Teori kepribadian yang terpopuler yang dikemukakan oleh Lewis Goldberg yaitu *Big Five Personality* yang membagi menjadi 5 yaitu *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, dan *Neuroticism* (Goldberg, R, 1992). Orang yang bersifat *openness* memiliki ciri lebih kreatif, terbuka terhadap hal baru yang ditemui, penasaran, dan berpikir luas. Orang yang bersifat *conscientiousness* memiliki ciri tekun dan dapat berorientasi pada pencapaian. Orang yang bersifat *extraversion* memiliki ciri mudah bersosialisasi dengan orang baru. Orang yang bersifat *agreeableness* memiliki penuh percaya diri, berhati lembut, dan suka menolong. Orang yang bersifat *neuroticism* memiliki emosi yang negatif sehingga rentan mengalami kecemasan, depresi dan sedih (Irawan, 2021).

Untuk mengetahui kepribadian seseorang harus mengikuti berbagai tes kepribadian. Tes tersebut dapat berupa deskripsi diri, observasi maupun wawancara yang dilakukan oleh ahli psikologi. Hal ini membuat kurang praktis dan memakan banyak waktu serta tempat. Ada beberapa tes kepribadian yang dilakukan pengguna di sosial media, namun tetap saja cara ini kurang praktis karena pengguna harus mengisi beberapa pertanyaan sehingga membutuhkan banyak waktu.

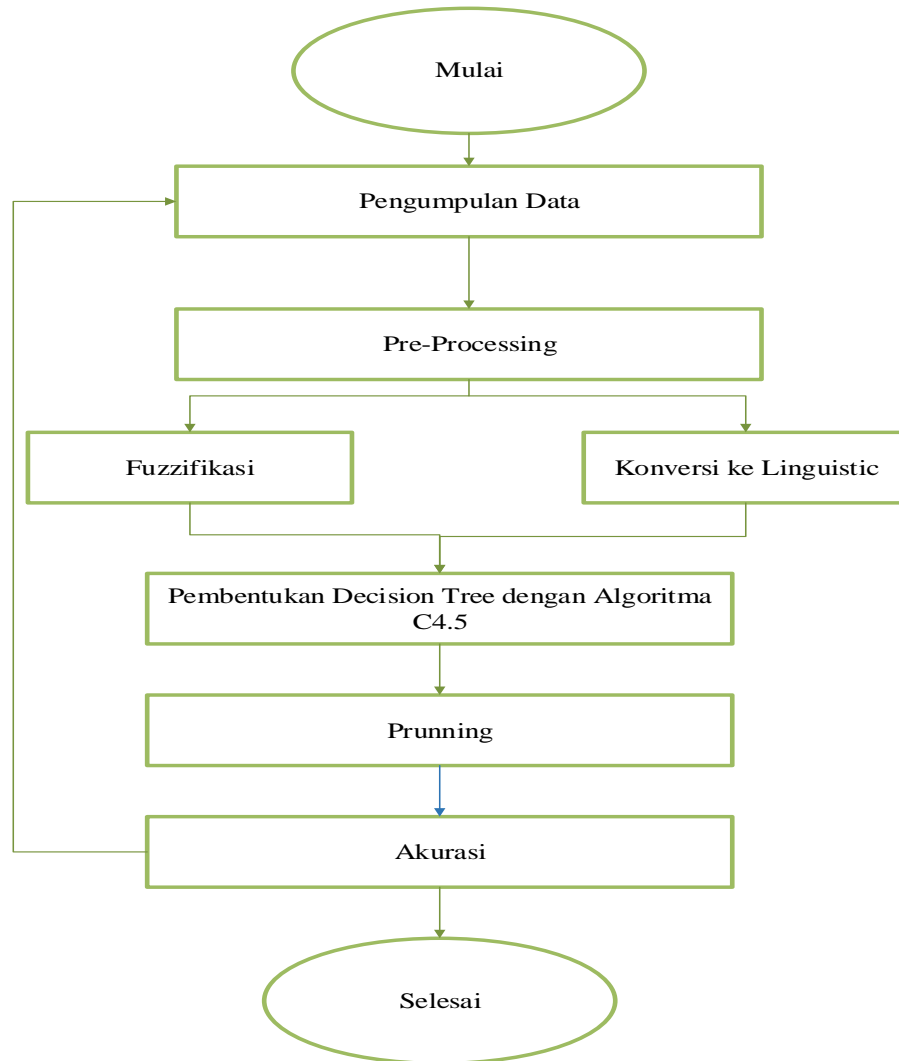
Banyak penelitian sudah dilakukan untuk memperdiksi kepribadian *Big Five* pada sosial media Twitter menggunakan *Term Frequency-Inverse Document Frequency* (TF - IDF) dan *Term Frequency-Relevance Frequency* (TF – RF) dengan metode *Naive Bayes* (Ilzam Nur Haq & Budi, 2019), C4.5 (Therik et al., 2021), dan *Logistic egression* (Ellandi et al., 2019). Pada penelitian yang menghasilkan percobaan menggunakan linguistik dengan pembobotan TF-IDF hampir memiliki akurasi yang sama dengan pembobotan TF-RF (Ilzam Nur Haq & Budi, 2019). Sedangkan dengan gabungan antara perilaku sosial dan linguistik meningkatkan nilai akurasi jika menggunakan salah satu dari pembobotan TF-IDF atau TF-RF. Sama halnya dengan penelitian yang menggunakan gabungan pendekatan linguistik dan perilaku sosial memiliki akurasi yang lebih tinggi dibandingkan hanya menggunakan pendekatan perilaku sosial (Zenico et al., 2019). Dengan menambahkan TF-IDF dan LIWC

dengan metode SMOTE untuk mendeteksi kepribadian *Big Five* perilaku sosial (Therik et al., 2021).

Pada penelitian ini klasifikasi kepribadian berdasarkan *Big Five Personality* sehingga menemukan kepribadian baru seperti *Extraverted, Serious, Dependable, Lively* dan *Responsible*. Data yang digunakan akan diekstraksi menggunakan metode *Fuzzy Decision Tree* dengan Algoritma C4.5. Model pembelajaran mesin *Decision Tree* menghasilkan model grafis yang dapat dinyatakan sebagai aturan. Pada proses induksi model *Decision Tree* sangat cepat dan dapat bersaing dengan hasil pembelajaran lainnya. Sistem *fuzzy* dapat diterapkan pada pembelajaran mesin *Decision Tree* (Cintra et al., 2013) dan dapat menangani ketidaktepatan melalui logika *fuzzy* dan teori himpunan *fuzzy* sehingga menghasilkan model yang dapat diinterpretasikan untuk mengklasifikasi kepribadian berdasarkan *Big Five Personality*. Algoritma C4.5 adalah algoritma yang paling relevan dan dapat digabungkan dengan *Fuzzy Decision Tree*. Untuk mendapatkan nilai ukuran dimensi-dimensi *Big Five* yang ada pada seseorang dengan cara menggunakan pendekatan leksikal yang memiliki unsur ketidakpastian. Untuk memecahkan masalah ketidakpastian dalam pendekatan leksikal diperlukan metode logika fuzzy (Mar'i et al., 2019), dan untuk mengolah dataset masukan yang berformat *numeric* pada tahap *pre-processing* menggunakan metode *fuzzy* C4.5 dimana jumlah *linguistic term* dari suatu atribut berpengaruh secara langsung dan signifikan terhadap akurasi sistem (Saptarini, 2016). Pada *Fuzzy Decision Tree* model yang paling kuat dari *Decision Tree* dengan interpretasi dan kemampuan untuk memproses sistem *fuzzy*. Gabungan dari sistem *fuzzy* dan *Decision Tree* telah menghasilkan model *Fuzzy Decision Tree* yang dapat menghasilkan model yang sederhana, akurat, dan dapat diinterpretasikan dengan biaya komputasi yang rendah. Pada penelitian lain tujuan penggunaan himpunan *fuzzy (linguistic term)* yang berbeda-beda pada klasifikasi kepribadian menggunakan algoritma *fuzzy* C4.5. Agar dapat mengetahui beberapa banyak jumlah himpunan *fuzzy (linguistic term)* yang ideal agar akurasi klasifikasi tinggi, dengan hasil percobaan 3 buah *linguistic term* pada 10 dataset memiliki akurasi 90% dan 182 memiliki akurasi 96% (Saptarini, 2016). Untuk klasifikasi kepribadian berdasarkan *Big Five Personality* menggunakan proses *pruning* untuk memotong *rule*. Dengan Penerapan metode *Fuzzy Decision Tree* dengan Algoritma C4.5 sehingga menghasilkan tingkat akurasi yang paling akurat pada klasifikasi kepribadian berdasarkan *Big Five Personality*.

METODE

Berikut adalah metode penelitian yang dilakukan untuk klasifikasi kepribadian berdasarkan *Big Five Personality*. Metode penelitian dapat dilihat pada gambar 1.



Gambar 1. Metode Penelitian

Dataset yang akan digunakan untuk penelitian ini terdapat 7 atribut yaitu *Gender*, *Age*, *Openness*, *Neuroticism*, *Conscientiousness*, *Agreeableness*, *Extraversion*, dan *Personality*. Data yang diperoleh merupakan open data yang didapat dalam bentuk excel.

Tahap Pre-processing merujuk pada penelitian (Saptarini, 2016) yang terdapat 2 proses utama pada tahap *pre-processing* yaitu proses fuzzifikasi untuk mengubah data numeric (*crisp*) menjadi *linguistic term* (*fuzzy*). Proses fuzzifikasi memiliki tujuan untuk

mengubah data yang bernilai numerik menjadi kabur (*fuzzy*) dengan menggunakan sebuah range yang akan ditentukan sebelumnya.

Pada tahapan pembentukan *Decision Tree* dengan algoritma C4.5 akan melakukan pengolahan data-data baru yang disebut dengan test dataset. Hasil dari pengolahan test dataset yang menggunakan algoritma C4.5 berupa pengelompokkan data ke dalam kelas-kelasnya. Pengelompokkan data akan membentuk suatu pohon keputusan. Pohon keputusan tersebut dibangun dengan cara membagi data secara rekursif hingga tiap bagian terdiri dari yang berasal dari kelas yang sama. Dalam penelitian ini terdapat 2 jenis *variable* data yang digunakan dalam klasifikasi kepribadian berdasarkan *Big Five Personality*, yaitu data kualitatif (nominal) dan data kuantitatif (numerik). *Variable* data kuantitatif berupa atribut dari *Big Five Personality* yaitu *Extraversion*, *Agreeableness*, *Conscientiousness*, *Neuroticism*, dan *Openness*. Dan untuk *variable* data kualitatif yaitu *Gender* dan *Age*.

Pada tahap selanjutnya yaitu teknik *pruning* bertujuan untuk mengontrol kompleksitas *decision tree* dan menggeneralisasikan *decision tree*. Ada dua macam metode dalam *pruning* dalam *decision tree*, yaitu *pre-pruning* yaitu menghentikan proses pembuatan cabang lebih awal dengan memutuskan untuk tidak lebih jauh mempartisi data training dan *post-pruning* yaitu memotong cabang yang kurang representatif setelah pohon selesai dibangun. *Pruning* juga dapat digunakan untuk mengatasi *overfitting* yang terjadi dikarenakan ada *noise* data *training* yaitu data yang tidak relevan sehingga mengakibatkan *decision tree* tidak seimbang (Faisal et al., 2017).

Tahap terakhir ini yaitu akurasi yang merupakan nilai derajat kedekatan dari pengukuran kuantitas untuk nilai sebenarnya (*true*). Untuk tahap ini menggunakan perhitungan akurasi dengan metode *Confusion Matrix* (Khamidah et al., 2018). Rumus dari akurasi sebagai berikut :

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

Pada penelitian ini dapat mengulangi kembali ke tahap pertama dengan menggunakan dataset kepribadian yang berisi 3 buah *linguistik term* jika *linguistik term* 2 sudah ditemukan akurasi.

HASIL DAN PEMBAHASAN

Data kepribadian *Big Five Personality Test* diambil dari *Repository Kaggle Machine Learning*. Data yang diperoleh akan dijadikan data latih dan data uji untuk model prediksi kepribadian berdasarkan *Big Five Personality*. Dataset ini diperoleh dalam bentuk excel dengan total 1026 record data dan 7 atribut yaitu *Gender, Openness, Neuroticism, Conscientiousness, Agreeableness, Extraversion*. Pelabelan pada dataset sudah tersedia pada *Repository Kaggle* sebanyak 227 baris untuk kelompok *Extraverted*, 314 baris untuk kelompok *Serious*, 159 baris untuk kelompok *Dependable*, 158 baris untuk kelompok *Lively*, dan 168 baris untuk kelompok *Responsible*. Akan dilakukan serangkaian proses pre-processing pada data dan kemudian disimpan kedalam format .CSV. Data yang diperoleh dapat dilihat pada tabel 1.

Tabel 1. Perolehan Data Kepribadian

No	Atribut
1	Gender
2	Age
3	Openness
4	Neuroticism
5	Conscientiousness
6	Agreeableness
7	Extraversion

Tahapan *Pre-processing* dilakukan untuk mengubah data agar bisa diolah oleh sistem dengan cara melengkapi data yang tidak lengkap, memasukkan nilai yang hilang dan mengakurasi data yang hilang.

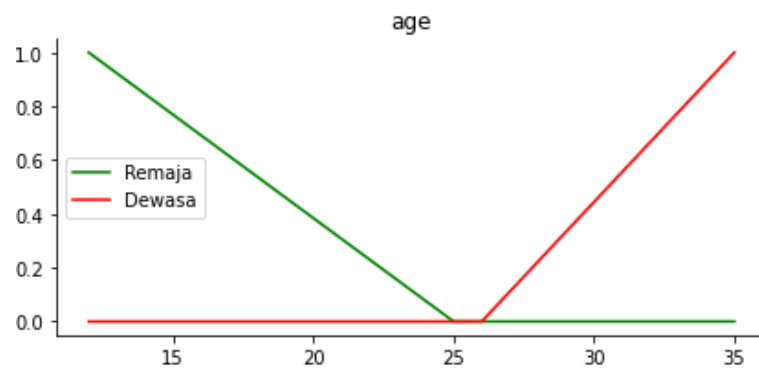
1. Atribut Age

Atribut age dibagi menjadi dua yaitu remaja ($12 \leq x < 26$) dan dewasa ($x < 36$). Dari pembagian tersebut dapat ditentukan *membership function* dari himpunan *fuzzy* remaja dan dewasa untuk atribut *age* secara terpisah yaitu :

$$\mu_{remaja}(x) = \begin{cases} 1; & x \leq 12 \\ \frac{25-x}{13} & 12 \leq x \leq 25 \\ 0; & x \geq 25 \end{cases}$$

$$\mu_{Dewasa}(x) = \begin{cases} 0 & x \leq 26 \\ \frac{x-26}{9} & 26 \leq x \leq 35 \\ 1; & x \geq 35 \end{cases}$$

Dalam himpunan *fuzzy* pada atribut age dapat digambarkan menggunakan *kurva* segitiga seperti gambar 2 dibawah.



Gambar 2. Transformasi Fuzzifikasi Atribut Age

2. Atribut Openness

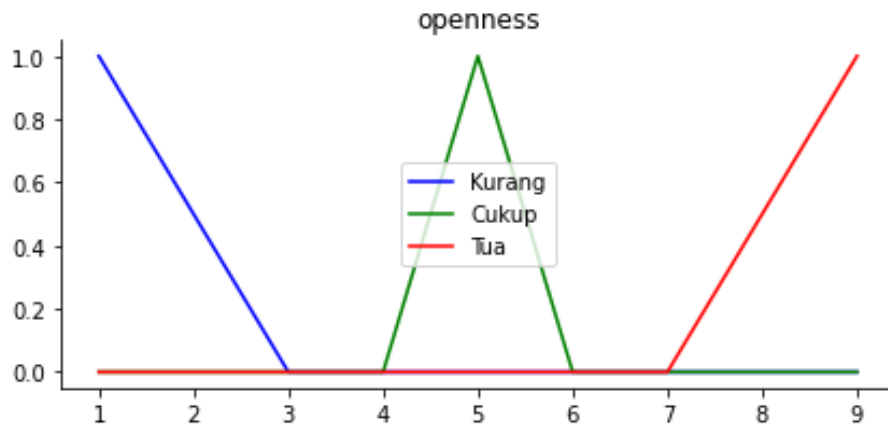
Atribut openness dibagi menjadi tiga yaitu kurang (1-3), cukup (4-6) dan sedang (7-9). Dari pembagian tersebut dapat ditentukan *membership function* dari himpunan *fuzzy* kurang, cukup dan sedang untuk atribut openness secara terpisah yaitu :

$$\mu_{kurang}(x) = \begin{cases} 1; & x \leq 1 \\ \frac{3-x}{2} & 1 \leq x \leq 3 \\ 0; & x \geq 3 \end{cases}$$

$$\mu_{Sedang}(x) = \begin{cases} 0 & x \leq 4 \text{ atau } x \geq 9 \\ \frac{x-4}{2} & 4 \leq x \leq 6 \\ \frac{9-x}{5} & 6 \leq x \leq 9 \end{cases}$$

$$\mu_{remaja}(x) = \begin{cases} 0; & x \leq 7 \\ \frac{x-7}{2} & 7 \leq x \leq 9 \\ 1; & x \geq 9 \end{cases}$$

Dalam himpunan *fuzzy* pada atribut openness dapat digambarkan menggunakan *kurva* segitiga seperti gambar 3 dibawah.



Gambar 3. Tranformasi Fuzzifikasi Atribut *Openness*

3. Atribut *Neuroticism*

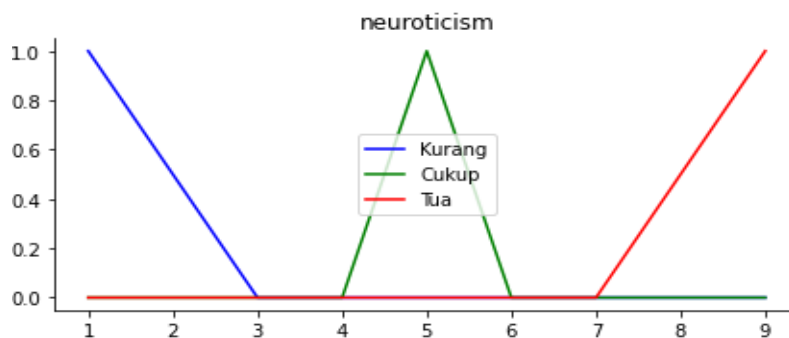
Atribut *neuroticism* dibagi menjadi tiga yaitu kurang (1-3), sedang (4-6) dan baik (7-9). Dari pembagian tersebut dapat ditentukan *membership function* dari himpunan *fuzzy* kurang, cukup dan sedang untuk atribut *neuroticism* secara terpisah yaitu :

$$\mu_{kurang}(x) = \begin{cases} 1; & x \leq 1 \\ \frac{3-x}{2} & 1 \leq x \leq 3 \\ 0; & x \geq 3 \end{cases}$$

$$\mu_{Sedang}(x) = \begin{cases} 0 & x \leq 4 \text{ atau } x \geq 9 \\ \frac{x-4}{2} & 4 \leq x \leq 6 \\ \frac{9-x}{2} & 6 \leq x \leq 9 \end{cases}$$

$$\mu_{remaja}(x) = \begin{cases} 0; & x \leq 7 \\ \frac{x-7}{2} & 7 \leq x \leq 9 \\ 1; & x \geq 9 \end{cases}$$

Dalam himpunan *fuzzy* pada atribut *neuroticism* dapat digambarkan menggunakan *kurva* segitiga seperti gambar 4 dibawah.



Gambar 4. Transformasi Fuzzifikasi Atribut *Neuroticism*

4. Atribut *Conscientiousness*

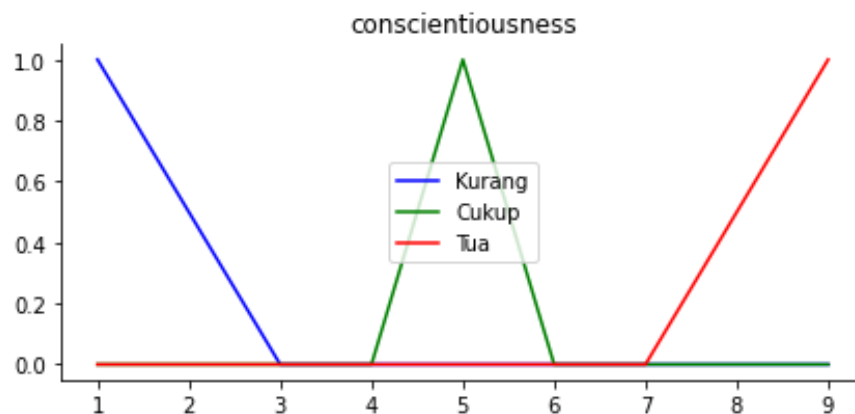
Atribut *conscientiousness* dibagi menjadi tiga yaitu kurang (1-3), cukup (4-6) dan baik (7-9). Dari pembagian tersebut dapat ditentukan *membership function* dari himpunan *fuzzy* kurang, cukup dan sedang untuk atribut *conscientiousness* secara terpisah yaitu :

$$\mu_{kurang}(x) = \begin{cases} 1; & x \leq 1 \\ \frac{3-x}{2} & 1 \leq x \leq 3 \\ 0; & x \geq 3 \end{cases}$$

$$\mu_{Sedang}(x) = \begin{cases} 0 & x \leq 4 \text{ atau } x \geq 9 \\ \frac{x-4}{2} & 4 \leq x \leq 6 \\ \frac{9-x}{5} & 6 \leq x \leq 9 \end{cases}$$

$$\mu_{remaja}(x) = \begin{cases} 0; & x \leq 7 \\ \frac{x-7}{2} & 7 \leq x \leq 9 \\ 1; & x \geq 9 \end{cases}$$

Dalam himpunan *fuzzy* pada atribut *conscientiousness* dapat digambarkan menggunakan *kurva* segitiga seperti gambar 5 dibawah.



Gambar 5. Transformasi Fuzzifikasi Atribut *Conscientiousness*

5. Atribut *Agreeableness*

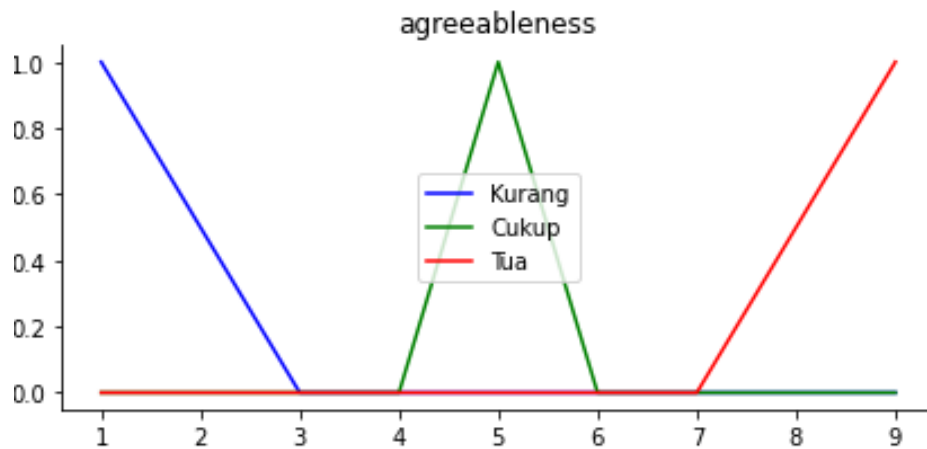
Atribut *agreeableness* dibagi menjadi tiga yaitu kurang (1-3), sedang (4-6) dan baik (7-9). Dari pembagian tersebut dapat ditentukan *membership function* dari himpunan *fuzzy* kurang, cukup dan sedang untuk atribut *agreeableness* secara terpisah yaitu

$$\mu_{kurang}(x) = \begin{cases} 1; & x \leq 1 \\ \frac{3-x}{2} & 1 \leq x \leq 3 \\ 0; & x \geq 3 \end{cases}$$

$$\mu_{Sedang}(x) = \begin{cases} 0 & x \leq 4 \text{ atau } x \geq 9 \\ \frac{x-4}{2} & 4 \leq x \leq 6 \\ \frac{9-x}{5} & 6 \leq x \leq 9 \end{cases}$$

$$\mu_{remaja}(x) = \begin{cases} 0; & x \leq 7 \\ \frac{x-7}{2} & 7 \leq x \leq 9 \\ 1; & x \geq 9 \end{cases}$$

Dalam himpunan *fuzzy* pada atribut *agreeableness* dapat digambarkan menggunakan *kurva* segitiga seperti gambar 6 dibawah.



Gambar 6. Transformasi Fuzzifikasi Atribut Agreeableness

6. Atribut *Extraversion*

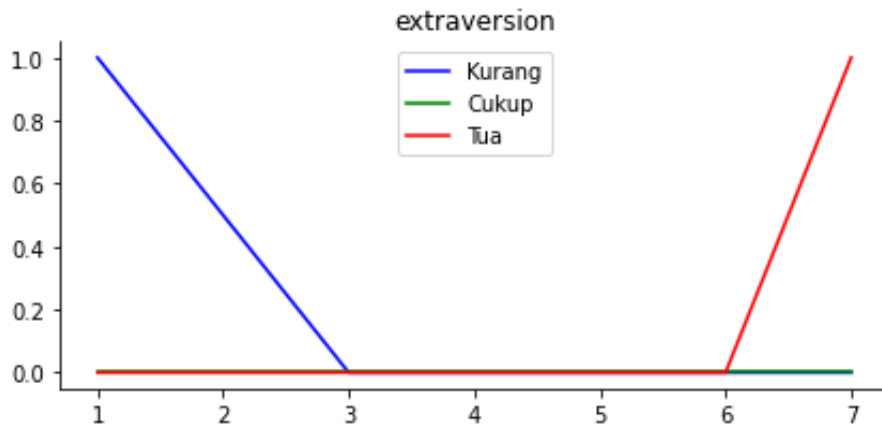
Atribut *extraversion* dibagi menjadi tiga yaitu kurang (1-3), sedang (4-5) dan baik (6-7). Dari pembagian tersebut dapat ditentukan *membership function* dari himpunan *fuzzy* kurang, cukup dan sedang untuk atribut *extraversion* secara terpisah yaitu :

$$\mu_{kurang}(x) = \begin{cases} 1; & x \leq 1 \\ \frac{3-x}{2} & 1 \leq x \leq 3 \\ 0; & x \geq 3 \end{cases}$$

$$\mu_{Sedang}(x) = \begin{cases} 0 & x \leq 4 \text{ atau } x \geq 7 \\ \frac{x-4}{1} & 4 \leq x \leq 5 \\ \frac{5-x}{2} & 5 \leq x \leq 7 \end{cases}$$

$$\mu_{remaja}(x) = \begin{cases} 0; & x \leq 6 \\ \frac{x-6}{1} & 6 \leq x \leq 7 \\ 1; & x \geq 7 \end{cases}$$

Dalam himpunan *fuzzy* pada atribut *extraversion* dapat digambarkan menggunakan *kurva* segitiga seperti gambar 8 dibawah.



Gambar 7. Tranformasi Fuzzifikasi Atribut *Extravesion*

Proses selanjutnya setelah tahapan *pre-processing* data. Data yang digunakan dalam pemodelan berjumlah 1.026 reocrd. Sebelum proses klasifikasi dengan menggunakan algoritma C4.5, data sebelumnya sudah dilakukan proses fuzzifikasi dan setiap atribut yang akan digunakan dikonversi ke linguistik. Berikut adalah tahapan-tahapan yang akan dilakukan pada pembentukan decision tree dengan algoritma C4.5.

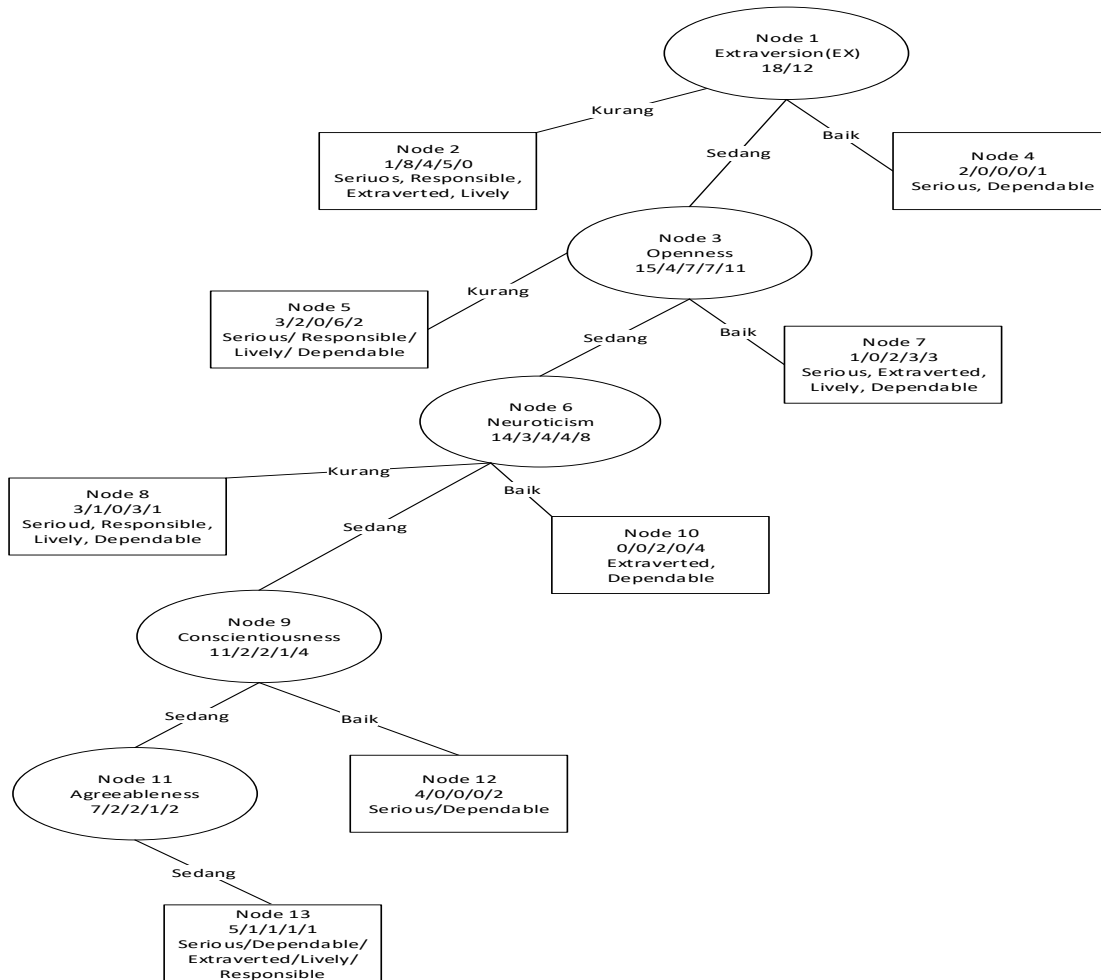
1. Menghitung *Entropy* dan *Gain*

Pada tahap pertama algoritma C4.5 ditentukan atribut yang akan digunakan sebagai *root* atau node akar. Untuk menentukan *root node* dengan cara menghitung *entropy* terlebih dahulu. Dari hasil perhitungan *entropy* akan digunakan untuk menghitung nilai *Gain* dari setiap atribut yang digunakan. Tabel 2 merupakan hasil dari perhitungan lengkap yang telah dilakukan pada setiap atribut.

Tabel 2. Perhitungan *Entropy & Gain*

	Jumlah kasus	Serious	Responsible	Extraverted	Lively	Dependable	Entropy	Gain
Total	66	18	12	12	12	12	2,299896	
Gender								0,123606
Female	35	7	10	7	3	8	2,235651	
Male	31	11	2	5	9	4	2,10927	
Age								0,56475
Remaja (12-25)	60	17	12	12	12	12	1,908661	
Dewasa (25-35)	1	1	0	0	0	0	0	
OP								0,546779
Kurang	6	1	4	1	0	0	0	
Sedang	51	16	8	9	9	9	2,26874	
Baik	9	1	0	2	3	3	0	
NE								0,200731
Kurang	21	5	4	0	9	3	0	
Sedang	35	13	8	6	3	5	2,158425	
Baik	10	0	0	6	0	4	0	
CO								0,439303
Kurang	10	1	3	3	0	3	0	
Sedang	37	9	7	7	10	4	2,262114	
Baik	19	8	2	2	2	5	2,057946	
AG								0,292081
Kurang	15	4	5	1	2	3	2,149255	
Sedang	44	13	7	8	9	7	2,279022	
Baik	7	1	0	3	1	2	0	
EX								0,841489
Kurang	18	1	8	4	5	0	0	
Sedang	44	15	4	7	7	11	2,187612	
Baik	4	2	0	0	0	1	0	

Setelah proses perhitungan gain ratio selesai, maka selanjutnya divisualisasikan pohon keputusan dari data yang digunakan saat perhitungan sebelumnya.



Gambar 8. Visualiasi Pohon Keputusan Tanpa Pruning

Setelah melakukan evaluasi untuk berbagai skenario yaitu pembagian data latih data uji dan menggunakan *pruning* dengan menggunakan linguistik 3 dan 5 dapat menghasilkan perbandingan akurasi seperti dapat dilihat pada tabel 3 dan tabel 4.

Tabel 3. Evaluasi Menggunakan 3 Linguistik

Pembagian Split Data	Dengan Pruning(%)	Tanpa Pruning(%)
60%40	47%	47%
70%30%	56%	56%
80%20%	59%	59%
90%10%	50%	50%

Tabel 4. Evaluasi Menggunakan 5 Linguistik

Pembagian Split Data	Dengan Pruning(%)	Tanpa Pruning(%)
60%405	57%	57%
70%30%	54%	56%
80%20%	48%	52%
90%10%	59%	53%

Pada tabel 3 dan tabel 4 diantara semua evaluasi skenario yang sudah dilakukan, skenario yang memiliki akurasi paling tinggi dengan dataset 3 linguistik pada pembagian split data 80:20 dengan tingkat akurasi sebesar 59% yang menggunakan *pruning* dan tanpa menggunakan *pruning*, skenario yang paling tinggi juga terdapat pada dataset 5 linguistik dengan pembagian split data 90:10 dengan menggunakan *pruning* memiliki tingkat akurasi sebesar 59% dan skenario yang memiliki akurasi yang paling rendah adalah pada dataset 3 linguistik dengan pembagian 60:40 memiliki tingkat akurasi sebesar 47% yang menggunakan *pruning* dan tidak menggunakan *pruning* dan pada dataset 5 linguistik dengan pembagian 80:20 memiliki akurasi sebesar 48% yang menggunakan *pruning*.

Evaluasi terhadap dataset 3 dan 5 linguistik dengan menggunakan *pruning* dan tanpa menggunakan *pruning* mempunyai tingkat akurasi dibawah 60% dikarenakan setelah mengubah dataset menjadi linguistik akan ada kemungkinan data tersebut sama dengan data yang lainnya sehingga akurasi yang didapat rendah. Akan tetapi dengan menggunakan 3

lingusik dan hanya menggunakan 2 kelas mendapatkan hasil yang bagus dengan menggunakan *pruning* memiliki tingkat akurasi 85% dan tanpa menggunakan *pruning* memiliki akurasi 74%.

KESIMPULAN DAN REKOMENDASI

Dalam penelitian Klasifikasi Kepribadian Berdasarkan *Big Five Personality* Menggunakan *Fuzzy Decision Tree* dengan Algoritma C4.5 menghasilkan sebuah model klasifikasi berdasarkan data kepribadian dengan atribut yang telah ditentukan berdasarkan penelitian-penelitian yang membahas tentang *Big Five Personality* yang mempengaruhi tingkat kepribadian seseorang. Hasil klasifikasi menggunakan algoritma C4.5 mendapatkan tingkat kurasi yang bervariasi berdasarkan rasio *split data* yang menggunakan *pruning* dan tanpa menggunakan *pruning* dengan linguistik 3 dan 5. Jika pada penelitian sebelumnya menggunakan *fuzzy decision tree* dengan algoritma C4.5 memiliki akurasi 64%, hasil akurasi pada penelitian ini memiliki akurasi paling tinggi dengan dataset 3 linguistik pada pembagian *split data* 80:20 dengan tingkat akurasi sebesar 59% yang menggunakan *pruning* dan tanpa menggunakan *pruning*. Skenario yang paling tinggi juga terdapat pada dataset 5 linguistik dengan pembagian *split data* 90:10 dengan menggunakan *pruning* memiliki tingkat akurasi sebesar 59%. Skenario yang memiliki akurasi yang paling rendah adalah pada dataset 3 linguistik dengan pembagian 60:40 memiliki tingkat akurasi sebesar 47% yang menggunakan *pruning* dan tidak menggunakan *pruning*. Pada dataset 5 linguistik dengan pembagian 80:20 memiliki akurasi sebesar 48% yang menggunakan *pruning*. Evaluasi terhadap dataset 3 dan 5 linguistik dengan menggunakan *pruning* dan tanpa menggunakan *pruning* mempunyai tingkat akurasi dibawah 60% dikarenakan setelah mengubah dataset menjadi linguistik akan ada kemungkinan data tersebut sama dengan data yang lainnya sehingga akurasi yang didapat rendah.

Pada penelitian ini direkomendasikan memilih atribut berdasarkan *Big Five Personality* yang mempengaruhi munculnya kepribadian lainnya. Atribut yang digunakan pada penelitian ini yaitu *Age*, *Openness*, *Neuroticism*, *Conscientiousness*, *Agreeableness*, dan *Extraversion*. Pada penelitian ini juga mempunyai kelas sebagai 5 kelas yaitu *Extraverted*, *Serious*, *Dependable*, *Lively*, dan *Responsible*. Pada dataset yang sudah melalui proses fuzzifikasi menggunakan 3 dan 5 linguistik dapat mempengaruhi tingkat akurasi yang

didapat. Sehingga pada penelitian selanjutnya kelas yang digunakan dapat menggunakan 2 dan 3 kelas saja agar mendapatkan akurasi yang paling bagus

REFERENSI

- Cintra, M., Monard, M., & Camargo, H. (2013). A Fuzzy Decision Tree Algorithm Based on C4.5. *Mathware & Soft Computing*, 20(1), 56–62.
- Ellandi, R., Budi, E., Si, S. S., Nugraha, F. N., & Psi, M. (2019). Prediksi kepribadian Big Five dengan Term-Frequency Inverse Document Frequency Menggunakan Metode k-Nearest Neighbor pada Twitter. *E-Proceeding of Engineering*, 6(2), 9955–9962.
- Faisal, M., Nasution, Y. N., & Ta, F. D. (2017). Perbandingan Kinerja Metode Klasifikasi Chi-square Automatic Interaction Detection (CHAID) dengan Metode Klasifikasi Algoritma C4 . 5 pada Studi Kasus : Penderita Diabetes Melitus Tipe 2 Di Samarinda Tahun 2015 Performance Comparison of Chi-square Autom. *Jurnal Eksponensial*, 8(2015), 119–124.
- Goldberg, R. L. (1992). The Development of Markers for the Big-Five Factor Structure. In *Psychological Assessment* (Vol. 4, Issue 1, pp. 26–42).
- Ilzam Nur Haq, F., & Budi, E. (2019). Implementasi Naive Bayes Classifier untuk Prediksi Kepribadian Big Five pada Twitter Menggunakan Term Frequency-Inverse Document Frequency (TF-IDF) dan Term Frequency-Relevance Frequency (TF-RF) Program Studi Sarjana Ilmu Komputasi Fakultas Informatik. *E-Proceeding of Engineering*, 6(2), 9785–9795.
- Irawan, A. S. Y. (2021). Prediksi Kepribadian Berdasarkan Media Sosial Twitter Menggunakan Metode Naïve Bayes Classifier. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 5(September), 988–996. <http://tunasbangsa.ac.id/ejurnal/index.php/jsakti/article/view/394>
- Khamidah, F. S. N., Hapsari, D. P., & Nugroho, H. (2018). Implementasi Fuzzy Decision Tree Untuk Prediksi Gagal Ginjal Kronis. *INTEGER: Journal of Information Technology*, 3(1), 19–28. <https://doi.org/10.31284/j.integer.2018.v3i1.155>

- Mar'i, F., Mahmudy, W. F., & Yusainy, C. (2019). Sistem Rekomendasi Profesi Berdasarkan Dimensi Big Five Personality Menggunakan Fuzzy Inference System Tsukamoto. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 6(5), 457. <https://doi.org/10.25126/jtiik.201965942>
- Saptarini, N. G. A. P. H. (2016). Penentuan Talenta Karyawan Berdasarkan Menggunakan Konsep Data Mining. *Jurnal Ilmiah Flash*, 2(1), 34. <https://doi.org/10.32511/jiflash.v2i1.22>
- Therik, S. V., Setiawan, E. B., & Telkom, U. (2021). *Deteksi Kepribadian Big Five Pengguna Twitter*. 8(5), 10277–10287.
- Yusup, A. H., Maharani, W., & Telkom, U. (2021). *Pembangunan Model Prediksi Kepribadian Berdasarkan Tweet Dan Kategori Kepribadian Big Five Dengan Metode Agglomerative*. 1(1), 44–50.
- Zenico, R., Setiawan, E. B., & Nugraha, F. N. (2019). Prediksi Big Five Personality dengan Term Frequency Inverse Document Frequency (TF – IDF) Menggunakan Metode Logistic Regression pada Pengguna Twitter. *E-Proceeding of Engineering Telkom University*, 6(2), 9939–9945.