

Klasifikasi Kepribadian Menggunakan Algoritma Decision Tree Berdasarkan Ten Item Personality Inventory

Carissa Limantara¹, Darsono Nababan²

^{1,2}Prodi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pelita Harapan,
Jl. Imam Bonjol No.6, Medan 20112
cl3095@student.uph.edu¹, darsono.nababan@uph.edu²

Abstrak – *The Ten Item Personality Inventory* (TIPI) dirancang sebagai alat pengukuran sederhana untuk menentukan lima kepribadian utama (*Big Five Personality Domain*) manusia: *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, dan *Neuroticism* (OCEAN). Dalam pengisian survey, responden cenderung merasa jenuh jika diminta untuk menjawab questionnaire yang panjang. Tes TIPI mengatasi hal tersebut dengan memberikan solusi sepuluh pertanyaan. Menggunakan algoritma decision tree, spreadsheet yang terdiri dari data responden tes TIPI akan diubah menjadi file xls, yang akan diproses melalui program data mining, *RapidMiner*. Kumpulan data yang dimuat akan dipecahkan menggunakan metode klasifikasi pohon keputusan, di mana sifat kepribadian yang dominan akan diterapkan sebagai atribut target. Hasil yang diharapkan dari studi ini adalah simulasi tes TIPI yang bisa menelaah kepribadian dominan responden secara akurat.

Kata Kunci – TIPI, OCEAN, data mining, pohon keputusan, RapidMiner

PENDAHULUAN

Ten Item Personality Inventory (TIPI), merupakan alat pengukur kepribadian sederhana yang dibuat untuk mengklasifikasikan kepribadian manusia ke dalam lima kategori (*The Big-Five*), diantaranya: *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, dan *Neuroticism* (OCEAN). Dengan menggunakan teknik data mining pohon keputusan, pengumpulan data kepribadian untuk tujuan pembelajaran dapat dipersingkat secara signifikan. Hasil akhir dari jurnal ini adalah sebuah simulasi untuk menentukan sifat dominan objek subjek pembelajaran dengan mudah.

Dalam pengumpulan data berbentuk questionnaire, tingkat pengabaian survei meningkat untuk survei yang memerlukan waktu lebih dari 7-8

menit; dengan tingkat penyelesaian menurun dari 5% hingga 20%. Selain itu, semakin banyak pertanyaan yang diajukan, semakin sedikit waktu yang dihabiskan responden untuk menjawab setiap pertanyaan [12]. Alhasil, kualitas dan keandalan data dapat menderita. Pada umumnya, jumlah pertanyaan tes psikologi yang banyak dapat memakan waktu yang Panjang, sehingga membuat jawaban responden kurang akurat. Misalnya tes psikologi Myers-Briggs Type Indicator (MBTI) yang berisi seratus pertanyaan. Jika diasumsikan rata-rata waktu penyelesaian satu soal adalah 15 detik, pengisi survey akan menghabiskan 25 menit untuk menyelesaikan tes MBTI.

Metode TIPI dapat mempersingkat survey menjadi sepuluh pertanyaan, sehingga jawaban yang didapatkan lebih akurat dan memungkinkan untuk ditambahkan pertanyaan lanjutan. Sampel data yang diperoleh dari openpsychometrics.org, *Artistic Preference Scale*, yang terdiri atas sepuluh atribut input (sepuluh variabel TIPI) dan satu atribut target (*dominant trait*).

A. The Big Five

Lima kepribadian utama yang didefinisikan oleh Digman (1990), Goldberg (1992), dan John & Srivastava (1999) adalah sebagai berikut:

1. *Openness* - Orang yang suka mempelajari hal-hal baru dan menikmati pengalaman baru biasanya mendapat nilai tinggi dalam keterbukaan. Keterbukaan termasuk sifat-sifat seperti berwawasan luas, imajinatif, dan memiliki berbagai kepentingan.
2. *Conscientiousness* - Orang yang memiliki tingkat kesadaran yang tinggi dapat diandalkan dan cepat. Ciri-ciri termasuk terorganisasi, metodis, dan menyeluruh.
3. *Extraversion* - Mendapatkan energi mereka dari berinteraksi dengan orang lain, sementara introvert mendapatkan energi mereka dari dalam diri mereka sendiri. Ekstraversi mencakup sifat-sifat energik, banyak bicara, dan tegas.

4. *Agreeableness* - Orang-orang ini ramah, kooperatif, dan penuh kasih sayang. Orang-orang dengan keramahan rendah mungkin lebih jauh. Ciri-cirinya termasuk bersikap baik, penyayang, dan simpatik.
5. *Neuroticism* - *Neuroticism* juga terkadang disebut Emotional Stability (kestabilan emosi). Dimensi ini berkaitan dengan stabilitas emosi dan tingkat emosi negatif seseorang. Orang yang mendapat skor tinggi pada neurotisme sering mengalami ketidakstabilan emosi dan emosi negatif. Ciri-cirinya termasuk moody dan tegang. [3] [7] [12]

B. Ten Item Personality Inventory (TIPI)

TIPI dirancang untuk menilai variasi kepribadian manusia yang ditentukan dari Teori Lima Faktor Kepribadian. Subjek penelitian akan diberikan sepuluh pernyataan yang akan diisi dengan angka satu sampai tujuh (1-7):

Berikut adalah peraturan khusus untuk menjumlahkan sepuluh komponen TIPI[12]:

1. *Item* yang mendapat skor terbalik adalah 2, 4, 6, 8, & 10. Contohnya mengodekan ulang 7 dengan 1, 6 dengan 2, 5 dengan 3, dst.
2. Ambil rata-rata dari dua *item* (*item* standar dan *item* yang diberi skor terbalik) yang membentuk setiap skala.

Tabel 1. Penilaian TIPI

Variabel	Arti Variabel	Rumus
TIPI1	Extraverted	E =
TIPI2	Critical	$\frac{TIPI1+TIPI2}{2}$
TIPI3	Dependable	A =
TIPI4	Anxious	$\frac{TIPI2+TIPI7}{2}$
TIPI5	Complex	C =
TIPI6	Quiet	$\frac{TIPI3+TIPI8}{2}$
TIPI7	Warm	N =
TIPI8	Careless	$\frac{TIPI4+TIPI9}{2}$
TIPI9	Calm	O =
TIPI10	Conventional	$\frac{TIPI5+TIPI10}{2}$

C. Data Mining

Data Mining mengacu pada penemuan pola baru dari banyak data dalam basis data dengan berfokus pada algoritma untuk menarik pengetahuan yang bermanfaat. [12]

Fayyad et.al. (1996) mendefinisikan enam fungsi utama *data mining* [2]:

1. Klasifikasi adalah menemukan model yang menganalisis dan menggolongkan data menjadi beberapa kelas yang telah ditentukan

2. Regresi adalah pemetaan *item* data menjadi variabel prediksi bernilai nyata
3. Clustering adalah mengidentifikasi serangkaian kategori atau *cluster* yang terbatas untuk menggambarkan data
4. Model Ketergantungan (*Association Rule Learning*) adalah mencari model yang menjelaskan dependensi signifikan antar variabel
5. Deteksi Deviasi (*Anomaly Detection*) adalah menemukan perubahan yang paling signifikan dalam data
6. *Summarization* adalah menemukan deskripsi ringkas untuk sub-kumpulan data

Gorunescu, F. (2011) menjelaskan bahwa *Data Mining* memiliki dua tujuan utama: prediksi dan deskripsi. [1]

Prediksi menggunakan beberapa variabel dalam kumpulan data untuk memprediksi nilai yang tidak diketahui dari variabel relevan lainnya (mis. klasifikasi, regresi, dan deteksi anomali). [1]

Deskripsi melibatkan pola data yang dimengerti manusia dan tren dalam data (misalnya pengelompokan, pembelajaran aturan asosiasi, dan *summarization*). [1]

D. Pohon Keputusan

Pohon keputusan merupakan salah satu model dari metode prediksi. Fungsi metode prediksi adalah untuk menentukan nilai yang hilang dari atribut target, dan memberikan prediksi dan penjelasan untuk prediksi tersebut. Nilai atribut hanya bisa diproduksi jika atribut target memiliki nilai yang perlu dicari[10].

Pohon keputusan berfungsi untuk melakukan klasifikasi nilai dari satu kolom berdasarkan nilai dari kolom lainnya. Cara kerja klasifikasi pohon keputusan adalah dengan mengenal pola atau dengan fungsi yang memisahkan kelas data yang satu dengan yang lainnya. Objek yang telah teridentifikasi dimasukkan ke dalam beberapa kategori berdasarkan kelakuan dan atribut dari kelompok yang telah terdefinisi. Klasifikasi hanya digunakan ketika nilai target bisa dikategorikan. Ini menjawab pertanyaan seperti, "akankah hasilnya menjadi A atau B?"[10]

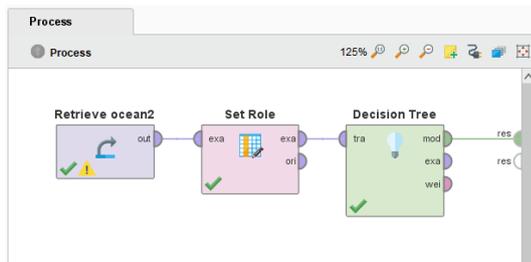
Jika atribut memadai, sangat mungkin untuk membangun sebuah pohon keputusan yang mengklasifikasikan setiap objek dalam *training set* dengan benar[4].

METODE PENELITIAN

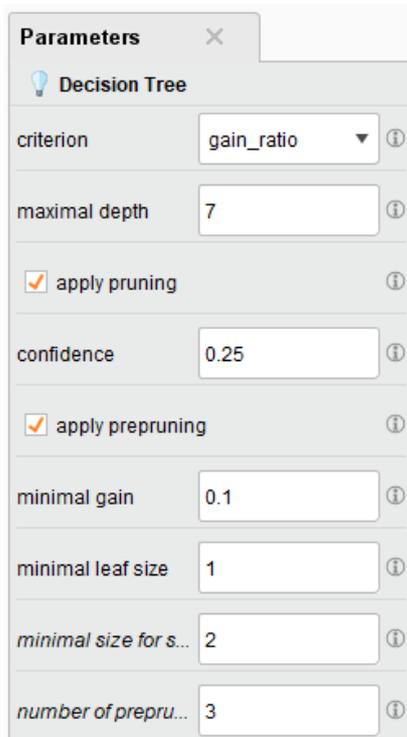
Metode yang digunakan untuk mensimulasikan proses penentuan kepribadian adalah metode prediksi pohon keputusan yang diproses melalui program *RapidMiner*. Sampel data diperoleh dari openpsychometrics.org, *Artistic Preference Scale*.

Berikut adalah langkah-langkah yang diambil untuk mendapatkan hasil data yang diproses menggunakan pohon keputusan:

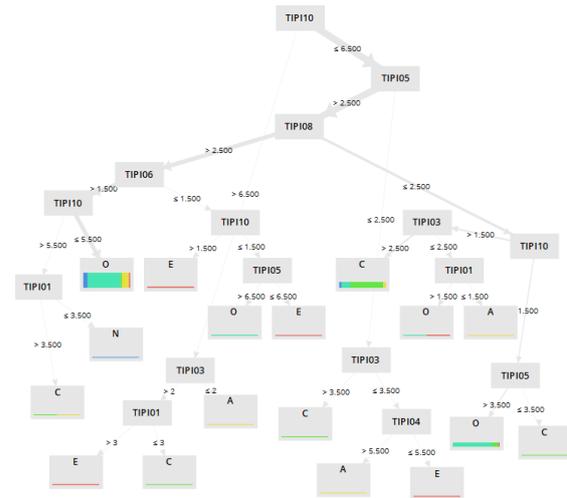
1. *Import data xml* ke dalam *Local Repository*, lalu dimuat ke *Process view*
2. Sambungkan proses ke operator pohon keputusan dan tentukan *role*
3. Terapkan parameter pohon keputusan
4. Eksekusi proses



Gambar 1. View Process



Gambar 2. Parameter Pohon Keputusan



Gambar 3. Grafik Pohon Keputusan

Tree

```

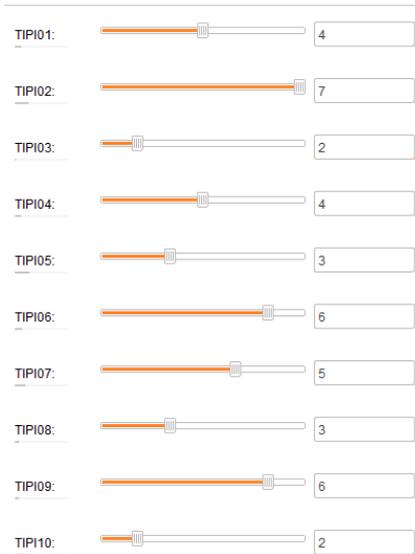
TIPI10 > 6.500
|
|_ TIPI03 > 2
|   |_ TIPI01 > 3: E {N=0, O=0, C=0, A=0, E=3}
|   |_ TIPI01 ≤ 3: C {N=0, O=0, C=2, A=0, E=0}
|   |_ TIPI03 ≤ 2: A {N=0, O=0, C=0, A=1, E=0}
|
|_ TIPI10 ≤ 6.500
|   |_ TIPI05 > 2.500
|   |   |_ TIPI08 > 2.500
|   |   |   |_ TIPI06 > 1.500
|   |   |   |   |_ TIPI10 > 5.500
|   |   |   |   |   |_ TIPI01 > 3.500: C {N=0, O=0, C=1, A=1, E=0}
|   |   |   |   |   |_ TIPI01 ≤ 3.500: N {N=1, O=0, C=0, A=0, E=0}
|   |   |   |   |   |_ TIPI10 ≤ 5.500: O {N=22, O=163, C=6, A=35, E=5}
|   |   |   |   |_ TIPI06 ≤ 1.500
|   |   |   |   |   |_ TIPI10 > 1.500: E {N=0, O=0, C=0, A=0, E=6}
|   |   |   |   |   |_ TIPI10 ≤ 1.500
|   |   |   |   |   |_ TIPI05 > 6.500: O {N=0, O=5, C=0, A=0, E=0}
|   |   |   |   |   |_ TIPI05 ≤ 6.500: E {N=0, O=0, C=0, A=0, E=1}
|   |   |   |_ TIPI08 ≤ 2.500
|   |   |   |   |_ TIPI10 > 1.500
|   |   |   |   |   |_ TIPI03 > 2.500: C {N=4, O=16, C=58, A=5, E=0}
|   |   |   |   |   |_ TIPI03 ≤ 2.500
|   |   |   |   |   |   |_ TIPI01 > 1.500: O {N=0, O=2, C=0, A=0, E=2}
|   |   |   |   |   |   |_ TIPI01 ≤ 1.500: A {N=0, O=0, C=0, A=1, E=0}
|   |   |   |   |_ TIPI10 ≤ 1.500
|   |   |   |   |   |_ TIPI05 > 3.500: O {N=0, O=47, C=8, A=0, E=1}
|   |   |   |   |   |_ TIPI05 ≤ 3.500: C {N=0, O=0, C=1, A=0, E=0}
|   |   |_ TIPI05 ≤ 2.500
|   |   |   |_ TIPI03 > 3.500: C {N=0, O=0, C=3, A=0, E=0}
|   |   |   |_ TIPI03 ≤ 3.500
|   |   |   |   |_ TIPI04 > 5.500: A {N=0, O=0, C=0, A=2, E=0}
|   |   |   |   |_ TIPI04 ≤ 5.500: E {N=0, O=0, C=0, A=0, E=2}
    
```

Gambar 4. Text View Pohon Keputusan

HASIL DAN PEMBAHASAN

Dengan menggunakan fitur *AutoModel* yang ada di *RapidMiner*, sebuah simulasi akan dihasilkan berdasarkan atribut *data set* yang dipakai. Berikut adalah contoh dan hasil yang diharapkan dari simulasi tes TIPI dengan metode pohon keputusan.

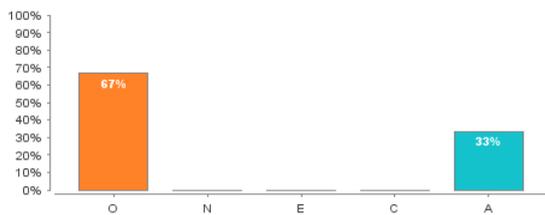
Input for Model



Gambar 4. Bagian *Input* Simulasi AutoModel

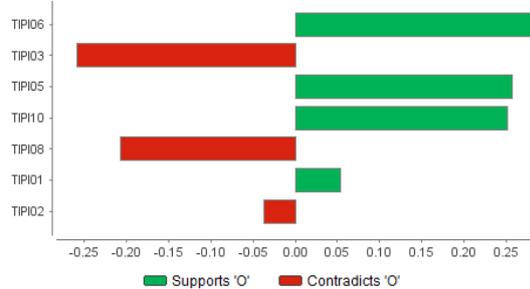
Berdasarkan data yang dimasukkan ke Gambar 4, terdapat tiga grafik pembuktian yang dihasilkan oleh AutoModel tersebut, berisikan kepribadian dominan yang dimiliki responden/subjek penelitian.

Prediction: O



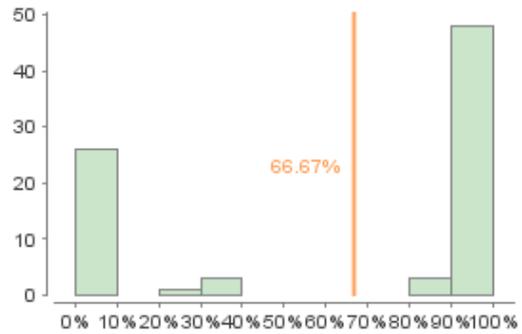
Gambar 5. Prediksi Kepribadian Dominan

Important Factors for O



Gambar 6. Variabel Penting N

Confidence Distribution for O



Gambar 7. Confidence Distribution

Important Factor menunjukkan *input* data yang mempengaruhi hasil prediksi serta perbandingannya dengan nilai-nilai lain, sebagaimana dipaparkan di gambar di atas.

Cox (1958) menyarankan bahwa distribusi kepercayaan "dapat didefinisikan secara langsung, atau dapat diperkenalkan dalam bentuk himpunan dengan semua interval kepercayaan pada berbagai tingkat probabilitas". Cox (1958) lebih lanjut menyatakan bahwa "Pernyataan dibuat atas dasar distribusi ini, asalkan kita berhati-hati tentang bentuk mereka, memiliki interpretasi frekuensi langsung."

Pada prinsipnya, menumpuk batas-batas satu set interval kepercayaan dari semua tingkat parameter biasanya dapat menghasilkan distribusi keyakinan untuk parameter-parameter lainnya. Dalam kasus ini, parameter tersebut adalah N.

KESIMPULAN DAN SARAN

1. Dengan menggunakan metode pohon keputusan disertai dengan implementasi program *RapidMiner*, proses pengolahan data TIPI menjadi lebih cepat
2. Metode pohon keputusan bisa digunakan untuk semua pengajian data survei yang berparameter pasti
3. Atribut pohon keputusan harus didefinisikan secara eksplisit
4. TIPI bisa diselesaikan dengan program *RapidMiner*
5. Kekurangan penggunaan metode *clustering* pohon keputusan adalah ketidakmampuan program untuk mengolah pertanyaan terbuka dan hanya terbatas pada pertanyaan tertutup

REFERENSI

- [1] F. Gorunescu. "Data Mining: Concepts, Models, and Techniques". India: Springer.
- [2] G. Piatetsky-Shapiro, P. Smyth, and U. Fayyad, "From Data Mining to Knowledge Discovery in Databases.," *AI Magazine*, vol. 17, no.3, pp. 37-54, 1996.
- [3] J. Digman, "Personality Structure: Emergence Of The 5-Factor Model," *Annual Review of Psychology*, vol. 41, no. 1, pp. 417-440, Jan. 1990.
- [4] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81-106, 1986.
- [5] K. Pousttchi and Y. Hufenbach, "Engineering the Value Network of the Customer Interface and Marketing in the Data-Rich Retail Environment," *International Journal of Electronic Commerce*, vol. 18, no. 4, pp. 17-42, 2014.
- [6] L., Morvant¹, Z., & Georges⁴, C. D. (n.d.). How much time are respondents willing to spend on your survey? Retrieved from https://www.surveymonkey.com/curiosity/survey_completion_times/
- [7] L. R. Goldberg, "The structure of phenotypic personality traits.," *American Psychologist*, vol. 48, no. 1, pp. 26-34, 1993.
- [8] M. -G. Xie and K. Singh, "Confidence Distribution, the Frequentist Distribution Estimator of a Parameter: A Review," *International Statistical Review*, vol. 81, no. 1, pp. 3-39, 2013.
- [9] Open psychology data: Raw data from online personality tests. [Online]. Available: https://openpsychometrics.org/_rawdata/. [Accessed: 26-Oct-2018].
- [10] RapidMiner Studio (Version 9.0) [Computer software]. (2018).
- [11] S. D. Gosling, P. J. Rentfrow, and W. B. Swann, "A very brief measure of the Big-Five personality domains," *Journal of Research in Personality*, vol. 37, no. 6, pp. 504-528, 2003.
- [12] T. Silwattananusarn, "Data Mining and Its Applications for Knowledge Management : A Literature Review from 2007 to 2012," *International Journal of Data Mining & Knowledge Management Process*, vol. 2, no. 5, pp. 13-24, 2012.