

# Prediksi Sifat Kepribadian Secara Tidak Langsung: Sebuah Studi Pustaka

Iman Paryudi

Program Studi Teknik Informatika  
Fakultas Teknik Universitas Pancasila, Jakarta, Indonesia  
iman.paryudi@univpancasila.ac.id

**Abstract** — Prediksi sifat kepribadian secara langsung mempunyai kelemahan karena dianggap membebani serta menyita waktu pengguna. Oleh karena itu, digunakan cara lain yaitu prediksi sifat kepribadian secara tidak langsung. Dua metoda prediksi sifat kepribadian secara tidak langsung adalah: (1) berdasarkan tulisan, (2) berdasarkan data demografi. Sebelum bisa memprediksi sifat kepribadian berdasarkan tulisan, harus dibuat terlebih dulu model yang menghubungkan antara sifat kepribadian dengan *linguistic cues*. Model ini dibuat dari corpus yang berisi tulisan dan sifat kepribadian. Sifat kepribadian bisa diprediksi dengan cara membandingkan *linguistic cues* seseorang yang diambil dari tulisannya di media sosial dengan model yang sudah dibuat. Seperti halnya pada prediksi sifat kepribadian berdasarkan tulisan, pada prediksi sifat kepribadian berdasarkan data demografi, model juga dibuat berdasarkan *corpus*. Bedanya, *corpus* disini berisi data demografi dan sifat kepribadian. Data ini kemudian dimodelkan untuk mendapatkan hubungan antara data demografi dan sifat kepribadian. Dari model ini, sifat kepribadian bisa diprediksi dengan cara membandingkan data demografi seseorang (misal umur dan jenis kelamin) dengan model yang sudah dibuat.

**Keywords:** *Prediksi sifat kepribadian secara langsung, Prediksi sifat kepribadian berdasarkan tulisan, Prediksi sifat kepribadian berdasarkan data demografi.*

## I. PENDAHULUAN

Beberapa tahun belakangan ini, penggunaan sifat kepribadian dalam sistem rekomendasi semakin meningkat. Hal ini karena adanya tiga keuntungan dari penggunaan sifat kepribadian dalam mengatasi cold-start problem, yaitu [1][2]: (i) Pengguna baru dapat segera mendapatkan rekomendasi yang akurat segera setelah menjadi anggota sistem rekomendasi tersebut [3], (ii) Sistem tidak perlu selalu menghitung kesamaan seorang pengguna dengan pengguna lainnya setelah pengguna tersebut memberikan rating pada sebuah item [4]. Hal ini karena profil pengguna berdasarkan sifat kepribadian tidak pernah berubah dan tidak dipengaruhi oleh data rating. (iii) Dalam hal sistem yang *cross-domain* (yaitu sebuah sistem yang merekomendasikan item dari beberapa domain), sistem semacam ini dapat menggunakan sifat kepribadian

yang sama untuk memberikan rekomendasi dari domain yang berbeda. Hal ini karena sifat kepribadian tidak terikat dengan domain tertentu [5].

Ada dua cara untuk memprediksi sifat kepribadian tidak langsung yaitu berdasarkan tulisan [6][7][8][9] dan berdasarkan data demografi [2]. Paper ini akan membahas mengenai bagaimana memprediksi sifat kepribadian berdasarkan tulisan dan data demografi. Bagian selanjutnya dari paper ini adalah deskripsi mengenai apa itu sifat kepribadian pada bagian 2. Bagian 3 dan 4 masing-masing membahas mengenai prediksi sifat kepribadian berdasarkan tulisan dan data demografi. Sedangkan ringkasan disajikan pada bagian 5.

## II. SIFAT KEPERIBADIAN

Sifat kepribadian didefinisikan sebagai kumpulan sifat yang dimiliki oleh seseorang yang mempengaruhi kognisi, emosi, motivasi, dan tingkah laku pada beberapa situasi [10]. Saat ini tes sifat kepribadian yang paling banyak digunakan adalah yang berdasarkan *big-five* [11]. Disebut *big five* karena ada lima faktor yang digunakan untuk menggolongkan sifat kepribadian manusia. Oleh karena itu metoda ini juga disebut dengan *Five Factor Model* (FFM).

*Big five* diciptakan berawal dari hipotesa yang menyatakan bahwa hal-hal yang paling penting dalam kehidupan manusia menjadi bagian dari bahasa mereka. Setelah mempelajari kegunaan dari bahasa, peneliti menemukan sekumpulan kata sifat yang menggambarkan sifat manusia. Melalui penelitian lanjutan, kumpulan kata sifat ini dikelompokkan lagi menjadi lima faktor utama yaitu *Openness to experience*, *Conscientiousness*, *Extraversion*, *Agreeableness*, dan *Neuroticism* yang disingkat dengan OCEAN [11]. Kata sifat-kata sifat yang menyusun setiap faktor adalah sebagai berikut:

*Openness* (O): artistik, ingin tahu, imajinatif, berwawasan luas, minat luas.

*Conscientiousness* (C): efisien, terorganisir, terencana, dapat diandalkan, bertanggung jawab, teliti.

*Extraversion* (E): aktif, tegas, enerjik, antusias, ramah, banyak bicara.

*Agreeableness* (A): menghargai, mudah memaafkan, dermawan, baik hati, simpatik, penuh kepercayaan.

*Neuroticism* (N): cemas, mengasihani diri sendiri, tegang, mudah tersinggung, tidak stabil, mudah khawatir.

*Openness* berhubungan dengan cara menginterpretasi situasi baru. Seseorang dengan skor *openness* tinggi menyukai hal-hal baru. Sedangkan seseorang dengan skor *openness* rendah menyukai sesuatu yang sudah dikenal dan diketahui. *Openness* juga mencakup derajat kecerdasan, keingintahuan, dan kreativitas seseorang bertemu dengan hal-hal baru.

*Conscientiousness* (sifat berhati-hati) berhubungan dengan cara mengorganisasi dan mengontrol tingkah laku atau hasrat. Seseorang dengan skor *conscientiousness* tinggi cenderung bijaksana, disiplin tinggi, taat, dan mampu mengontrol hasrat. Sedangkan seseorang dengan skor *conscientiousness* rendah cenderung untuk bertindak menurut kata hati (*impulsif*), tidak mempunyai tujuan jelas, serta tidak terorganisasi dengan baik.

*Extraversion* berhubungan dengan derajat keinginan untuk berhubungan dengan dunia luar. Seseorang dengan skor *extraversion* tinggi (*extrovert*) suka berhubungan dengan dunia luar dan enerjik. Sedangkan seseorang dengan skor *extraversion* rendah (*introvert*) cenderung pendiam dan tidak suka bersosialisasi.

*Agreeableness* menunjukkan bagaimana empati kita kepada orang lain. Seseorang dengan skor *agreeableness* tinggi adalah orang yang ramah, suka menolong, dan kooperatif. Sedangkan seseorang dengan skor *agreeableness* rendah adalah orang yang selalu curiga, tidak mudah percaya pada orang lain, dan cenderung untuk mementingkan diri sendiri.

*Neuroticism* mengukur kestabilan emosi seseorang. Seseorang dengan skor *neuroticism* tinggi adalah orang yang suka marah, mudah cemas, dan suka murung. Sedangkan seseorang dengan skor *neuroticism* rendah adalah orang yang emosinya stabil.

Terdapat banyak macam kuesioner yang biasa digunakan untuk memprediksi sifat kepribadian berdasarkan big five dimana perbedaannya adalah jumlah pertanyaan pada kuesioner. Kuesioner yang terpanjang adalah *Global Personality Inventory* (GPI) dengan 504 pertanyaan dan yang terpendek adalah *Ten-Item Personality Inventory* (TIPI) dengan hanya 10 pertanyaan [12][3][13]. Berbagai macam kuesioner sudah dipergunakan oleh beberapa peneliti dalam penelitian mereka, diantaranya adalah: *Big Five Inventory* (BFI) yang berisi 44 pertanyaan [14][15][16][17], *International Personality Item Pool* (IPIP) yang berisi 50 pertanyaan [2][18][4], *Big Five Markers* (BFM) [10], *NEO-IPIP* yang berisi 300 pertanyaan [12], *Revised NEO Personality Inventory* (NEO-PI-R) yang berisi 240 pertanyaan [19]. bagian 5.

### III. PREDIKSI SIFAT KEPERIBADIAN BERDASARKAN TULISAN

Apabila prediksi sifat kepribadian secara langsung dilakukan dengan menggunakan kuesioner, maka pada prediksi sifat kepribadian secara tidak langsung dilakukan dengan menghubungkan antara apa yang dimiliki oleh seseorang dengan sifat kepribadiannya. Pada prediksi sifat kepribadian secara tidak langsung berdasarkan tulisan, yang dicari adalah hubungan antara gaya penulisan dari seseorang dengan sifat kepribadiannya. Tulisan yang dipakai pada prediksi ini adalah tulisan yang ditulis di media sosial seperti Weblog [18], Twitter [15][17][19], Facebook [16][20][21][19][22][23], FriendFeed [24], TripAdvisor [25][10], serta Instagram [17].

Yang paling penting pada prediksi sifat kepribadian secara tidak langsung berdasarkan tulisan adalah *corpus*. *Corpus* digunakan untuk memprediksi sifat kepribadian pengguna. Celli [24] mengumpulkan data untuk *corpus* nya dari 748 pengguna FriendFeed dengan 1065 postingan. Sementara itu, *corpus* Weblog dikumpulkan oleh [18] dari 71 responden (47 wanita dan 24 pria) dengan meminta mereka untuk menulis mengenai studi mereka di blog mereka masing-masing. Mereka juga diminta untuk menjawab kuesioner sifat kepribadian menggunakan IPIP50. *Corpus* inilah yang akan digunakan untuk membuat model hubungan antara gaya tulisan (berupa kata-kata yang sering dipakai oleh pengguna atau *linguistic cues*) dengan sifat kepribadian.

Proses untuk memprediksi sifat kepribadian berdasarkan tulisan adalah sebagai berikut:

- Pengumpulan tulisan. Tulisan dikumpulkan dari media sosial. Supaya bisa diproses lebih lanjut, maka tulisan tersebut harus memenuhi kriteria panjang tertentu. Misalnya, [16] menggunakan batas panjang tulisan rata-rata 42.6 kata.
- Pemrosesan tulisan. Pada tahap ini, dilakukan ekstraksi *linguistic cues*. *Linguistic cues* adalah kata-kata yang paling sering digunakan oleh pengguna. Untuk dapat mengekstraksi *linguistic cues*, tahapannya adalah [10][15]: tulisan dikonversi ke huruf kecil, tokenisasi, menghapus tanda henti kata seperti titik dan koma, *stemming*, *text tagging*, dan menghapus tulisan yang terlalu pendek. Tulisan yang sudah diproses ini kemudian dikirim ke tool untuk menganalisa tulisan seperti LIWC untuk mendapatkan *linguistic cues* [25][16].
- Prediksi sifat kepribadian. Dengan menggunakan *linguistic cues* yang didapatkan pada bagian sebelumnya, selanjutnya *linguistic cues* ini dibandingkan dengan model hubungan antara sifat kepribadian dengan *linguistic cues* yang didapat dari *corpus*.

#### IV. PREDIKSI SIFAT KEPRIIBADIAN BERDASARKAN DATA DEMOGRAFI

Kalau pada prediksi sifat kepribadian berdasarkan tulisan, model dibuat dari corpus, maka disini model dibuat sepenuhnya berdasarkan survei. Dalam survei tersebut, responden diminta untuk mengisi kuesioner yang berisi pertanyaan mengenai data demografi (seperti tahun lahir, jenis kelamin, dsb.) serta pertanyaan mengenai sifat kepribadian. Dalam kasus ini, [2] menggunakan kuesioner sifat kepribadian IPIP50.

Setelah mendapatkan data, dilakukan pemrosesan terhadap data tersebut:

- Mengubah tahun lahir menjadi umur
- Menghitung skor total sifat kepribadian masing-masing responden.
- Berdasarkan skor sifat kepribadian yang didapat, dilakukan penggolongan sifat kepribadian menjadi tinggi dan rendah.
- Berdasarkan umur yang didapat, dilakukan pengelompokan umur menjadi remaja, dewasa, dan lansia.
- Memilih atribut yang akan digunakan dalam pemodelan: golongan darah, pekerjaan, warna kesukaan, jenis kelamin, hobi, olah raga kesukaan, zodiak, kelompok umur, status perkawinan, suku/etnis, tingkat sifat kepribadian.

Tahap selanjutnya adalah membuat model. Dalam pembuatan model, [2] menggunakan metoda Decision Tree dan evaluasi model menggunakan *10-fold Cross-Validation*. Dalam hal ini data demografi dijadikan sebagai *independent attribute*, sedangkan tingkat sifat kepribadian dijadikan sebagai *dependent attribute*. Dari pemodelan ini didapatkan bahwa yang data demografi yang paling berpengaruh pada sifat kepribadian adalah umur dan jenis kelamin. Model yang didapatkan dari tahap ini, selanjutnya digunakan untuk memprediksi sifat kepribadian seseorang. Misal seseorang dengan kelompok umur tertentu (misal remaja) dan jenis kelamin tertentu (misal wanita), maka akan mempunyai sifat kepribadian tertentu.

#### V. RINGKASAN

Selain bisa diprediksi secara langsung menggunakan kuesioner, sifat kepribadian juga bisa diprediksi secara tidak langsung yaitu berdasarkan tulisan dan data demografi milik seseorang. Prediksi sifat kepribadian berdasarkan tulisan dilakukan dengan membandingkan *linguistic cues* dari seseorang dengan model sifat kepribadian yang menghubungkan antara *linguistic cues* dengan sifat kepribadian. Sementara itu, prediksi sifat kepribadian berdasarkan data demografi dilakukan dengan terlebih dulu membuat model yang menghubungkan antara data demografi dengan sifat kepribadian. Dari model ini, sifat kepribadian seseorang

bisa diprediksi hanya dari data demografi dari orang tersebut.

#### DAFTAR PUSTAKA

- [1] I. Paryudi, A. Ashari, and A. M. Tjoa, "Personality Estimation using Demographic Data in a Personality-based Recommender System: A Proposal," 2019, doi: 10.1145/3366030.3366098.
- [2] I. Paryudi, A. Ashari, and K. Mustofa, "The Performance of Personality-based Recommender System for Fashion with Demographic Data-based Personality Prediction," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 1, pp. 360–368, 2022, doi: 10.14569/IJACSA.2022.0130145.
- [3] R. Hu and P. Pu, "Enhancing Collaborative Filtering Systems with Personality Information," in *RecSys '11*, 2011, pp. 197–204.
- [4] M. Tkalčić, M. Kunaver, J. Tasic, and A. Kosir, "Personality based user similarity measure for a collaborative recommender system," 2009. [Online]. Available: <https://www.researchgate.net/publication/228984904>.
- [5] M. Tkalčić and L. Chen, "Personality and Recommender Systems," in *Recommender Systems Handbook*, Second Edi., F. Ricci, L. Rokach, and B. Shapira, Eds. New York: Springer, 2015, pp. 715–739.
- [6] I. Paryudi and S. R. C. Nursari, "Graph-based Personality Modeling for Personality-based Recommender System : Do Race and Questionnaire Type Affect Model Testing Result ?," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 3, pp. 3427–3440, 2020.
- [7] I. Paryudi, E. Winarko, S. Priyanta, and S. R. Candra Nursari, "Modeling of Personality Traits based on Demographic Data on Multi-Races Samples of Ages from 13 to 50 Years Old: Investigating the Effect of Race on Model," in *2020 6th International Conference on Science and Technology (ICST)*, 2020, pp. 1–6, doi: 10.1109/ICST50505.2020.9732792.
- [8] I. Paryudi, A. Ashari, and K. Mustofa, "Modelling Relationship between Demographic Data and Personality Traits," *Turkish J. Comput. Mat. Educ.*, vol. 12, no. 14, pp. 2247–2255, 2021.
- [9] I. Paryudi, A. Ashari, and K. Mustofa, "Creating Personality and Preference Models based on Demographic Data for Personality-based Recommender System for Fashion using Decision Tree and Association Rule," *Turkish J. Comput. Mat. Educ.*, vol. 12, no. 14, pp. 5165–5174, 2021, [Online]. Available: <https://turcomat.org/index.php/turkbilmat/article/view/11548>.
- [10] A. Di Rienzo and A. Neishabouri, "Recommendations with personality traits extracted

- from text reviews,” in *Studies in Computational Intelligence*, vol. 616, Springer Verlag, 2016, pp. 355–364.
- [11] R. R. McCrae and O. P. John, “An Introduction to the Five-Factor Model and Its Applications,” *J. Pers.*, vol. 60, no. 2, pp. 175–215, 1992, doi: <https://doi.org/10.1111/j.1467-6494.1992.tb00970.x>.
- [12] M. Augusta Silveira Netto Nunes, “Recommender Systems based on Personality Traits,” Université Montpellier II, 2008.
- [13] F. Lu and N. Tintarev, “A Diversity Adjusting Strategy with Personality for Music Recommendation,” 2018.
- [14] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore, “Using linguistic cues for the automatic recognition of personality in conversation and text,” *J. Artif. Intell. Res.*, vol. 30, pp. 457–500, 2007, doi: [10.1613/jair.2349](https://doi.org/10.1613/jair.2349).
- [15] J. Golbeck, C. Robles, M. Edmondson, and K. Turner, “Predicting personality from twitter,” *Proc. - 2011 IEEE Int. Conf. Privacy, Secur. Risk Trust IEEE Int. Conf. Soc. Comput. PASSAT/SocialCom 2011*, pp. 149–156, 2011, doi: [10.1109/PASSAT/SocialCom.2011.33](https://doi.org/10.1109/PASSAT/SocialCom.2011.33).
- [16] J. Golbeck, C. Robles, and K. Turner, “Predicting personality with social media,” in *Conference on Human Factors in Computing Systems - Proceedings*, 2011, pp. 253–262, doi: [10.1145/1979742.1979614](https://doi.org/10.1145/1979742.1979614).
- [17] M. Skowron, M. Tkalčič, B. Ferwerda, and M. Schedl, “Fusing Social Media Cues: Personality Prediction from Twitter and Instagram,” in *25th International Conference Companion on World Wide Web - WWW '16 Companion*, 2016, pp. 107–108, doi: [10.475/123](https://doi.org/10.475/123).
- [18] J. Oberlander and S. Nowson, “Whose thumb is it anyway? Classifying author personality from weblog text,” in *Proceedings of the COLING/ACL*, Dec. 2006, pp. 627–634.
- [19] G. Carducci, G. Rizzo, D. Monti, E. Palumbo, and M. Morisio, “TwitPersonality: Computing personality traits from tweets using word embeddings and supervised learning,” *Inf.*, vol. 9, no. 5, pp. 1–20, 2018, doi: [10.3390/info9050127](https://doi.org/10.3390/info9050127).
- [20] D. Markovikj, S. Gievska, M. Kosinski, and D. Stillwell, “Mining facebook data for predictive personality modeling,” *AAAI Work. - Tech. Rep.*, vol. WS-13-01, pp. 23–26, 2013.
- [21] K. H. Peng, L. H. Liou, C. S. Chang, and D. S. Lee, “Predicting personality traits of Chinese users based on Facebook wall posts,” in *2015 24th Wireless and Optical Communication Conference, WOCC 2015*, 2015, pp. 9–14, doi: [10.1109/WOCC.2015.7346106](https://doi.org/10.1109/WOCC.2015.7346106).
- [22] Z. Wang, C. H. Wu, Q. B. Li, B. Yan, and K. F. Zheng, “Encoding text information with graph convolutional networks for personality recognition,” *Appl. Sci.*, vol. 10, no. 12, 2020, doi: [10.3390/APP10124081](https://doi.org/10.3390/APP10124081).
- [23] X. Wang, Y. Sui, K. Zheng, Y. Shi, and S. Cao, “Personality classification of social users based on feature fusion,” *Sensors*, vol. 21, no. 20, 2021, doi: [10.3390/s21206758](https://doi.org/10.3390/s21206758).
- [24] F. Celli, “Unsupervised Personality Recognition for Social Network Sites,” *ICDS 2012, Sixth Int. Conf. Digit. Soc.*, no. c, pp. 59–62, 2012.
- [25] A. Roshchina, J. Cardiff, and P. Rosso, “TWIN: Personality-based Intelligent Recommender System,” *J. Intell. Fuzzy Syst.*, vol. 28, no. 5, pp. 2059–2071, 2015, doi: [10.3233/IFS-141484](https://doi.org/10.3233/IFS-141484).