

**Big Five Personality Assessment Using KNN method with RoBERTA**Athirah Rifdha Aryani¹, Erwin Budi Setiawan²^{1,2}Fakultas Informatika, Universitas Telkom, Bandung¹athirahrifdha@students.telkomuniversity.ac.id, ²erwinbudisetiawan@telkomuniversity.ac.id**Reference**

- [1] N. Febrianto, I. Prasetya, and A. Wijaya, "Pembuatan Sistem Prediksi Kepribadian 'The Big Five Traits' dari Media Sosial Twitter." [Online]. Available: http://semioacast.com/en/publications/2012_07_30_Twitter_reaches_half_a_billion_
- [2] M. G. Tambunan¹ and E. B. Setiawan, "Prediksi Kepribadian DISC Pada Twitter Menggunakan Metode Decision Tree C4.5 dengan Pembobotan TF-IDF dan TF-RF."
- [3] R. Ellandi, E. Budi, S. S. Si, N. Fida, S. Nugraha, and M. P. Psi, "Prediksi kepribadian Big Five dengan Term-Frequency Inverse Document Frequency Menggunakan Metode k-Nearest Neighbor pada Twitter."
- [4] G. D. Salsabila and E. B. Setiawan, "Semantic Approach for Big Five Personality Prediction on Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 680–687, Aug. 2021, doi: 10.29207/resti.v5i4.3197.
- [5] Y. Liu *et al.*, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," Jul. 2019, [Online]. Available: <http://arxiv.org/abs/1907.11692>
- [6] F. Celli and B. Lepri, "Is Big Five better than MBTI? A personality computing challenge using Twitter data." [Online]. Available: <https://twitter.com/search-advanced>
- [7] C. Yuan, J. Wu, H. Li, and L. Wang, *Personality Recognition based on User Generated Content*. IEEE, 2018.
- [8] J. Eka Sembodo, E. Budi Setiawan, and Z. Abdurahman Baizal, "Data Crawling Otomatis pada Twitter," Sep. 2016, pp. 11–16. doi: 10.21108/indosc.2016.111.
- [9] "Big Five Personality Test." <https://bigfive-test.com/> (accessed Jul. 09, 2022).
- [10] B. Yudha Pratama NRP, A. Ec Ir Rianarto Sarno, and R. A. Nur Esti, "Personality Classification Based on Twitter Text Using Naive Bayes, KNN and SVM."
- [11] L. Peterson, "K-nearest neighbor," *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009, doi: 10.4249/scholarpedia.1883.
- [12] D. Faraj and M. Abdullah, "SarcasmDet at SemEval-2021 Task 7: Detect Humor and Offensive based on Demographic Factors using RoBERTa Pre-trained Model."
- [13] S. Lei, "A feature selection method based on information gain and genetic algorithm," in *Proceedings - 2012 International Conference on Computer Science and Electronics Engineering, ICCSEE 2012*, 2012, vol. 2, pp. 355–358. doi: 10.1109/ICCSEE.2012.97.
- [14] F. Last, G. Douzas, and F. Bacao, "Oversampling for Imbalanced Learning Based on K-Means and SMOTE," Nov. 2017, doi: 10.1016/j.ins.2018.06.056.
- [15] H. Hairani, K. E. Saputro, and S. Fadli, "K-means-SMOTE for handling class imbalance in the classification of diabetes with C4.5, SVM, and naive Bayes," *Jurnal Teknologi dan Sistem*

- Komputer*, vol. 8, no. 2, pp. 89–93, Apr. 2020, doi: 10.14710/jtsiskom.8.2.2020.89-93.
- [16] R. Ghawi and J. Pfeffer, “Efficient Hyperparameter Tuning with Grid Search for Text Categorization using kNN Approach with BM25 Similarity,” *Open Computer Science*, vol. 9, no. 1, pp. 160–180, Jan. 2019, doi: 10.1515/comp-2019-0011.
- [17] Willy, E. B. Setiawan, and F. N. Nugraha, “Implementation of Decision Tree C4.5 for Big Five Personality Predictions with TF-RF and TF-CHI2 on Social Media Twitter,” in *2019 International Conference on Computer, Control, Informatics, and its Applications: Emerging Trends in Big Data and Artificial Intelligence, IC3INA 2019*, Oct. 2019, pp. 114–119. doi:10.1109/IC3INA48034.2019.8949601.
- [18] K. Prameswari and E. B. Setiawan, “Analisis Kepribadian Melalui Twitter Menggunakan Metode Logistic Regression dengan Pembobotan TF-IDF dan AHP.”