UNIVERSITAS
GADJAH MADA
**Deteksi Code Smell Menggunakan Kombinasi Metrik Struktural dan Representasi Semantik Kode pada
Dataset yang Relevan dengan Industri**
Azhar Bagaskara, Arif Nurwidyantoro, M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR PUSTAKA

Abdi, H. dan Williams, L. J., July 2010, Principal component analysis, *WIREs Computational Statistics*, 4, 2, 433–459.

Alexandropoulos, S.-A. N., Kotsiantis, S. B., dan Vrahatis, M. N., 2019, Data preprocessing in predictive data mining, *The Knowledge Engineering Review*, 34, e1.

Ali, I., Rizvi, S. S. H., dan Adil, S. H., Apr. 21, 2025, Enhancing Software Quality with AI: A Transformer-Based Approach for Code Smell Detection, *Applied Sciences*, 8, 15, 4559.

Alon, U., Zilberstein, M., Levy, O., dan Yahav, E., *code2vec: Learning Distributed Representations of Code*, Oct. 30, 2018, DOI: 10.48550/arXiv.1803.09473, arXiv: 1803.09473[cs], *http://arxiv.org/abs/1803.09473* (diakses pada 11/28/2025).

Arcelli Fontana, F., Mäntylä, M. V., Zanoni, M., dan Marino, A., June 2016, Comparing and experimenting machine learning techniques for code smell detection, *Empirical Software Engineering*, 3, 21, 1143–1191.

Arif, A. dan Rana, Z. A., Dec. 16, 2020, Refactoring of Code to Remove Technical Debt and Reduce Maintenance Effort, *2020 14th International Conference on Open Source Systems and Technologies (ICOSST)*, Lahore, Pakistan, 1–7.

Bhave, A. dan Sinha, R., Oct. 2022, Deep Multimodal Architecture for Detection of Long Parameter List and Switch Statements using DistilBERT, *2022 IEEE 22nd International Working Conference on Source Code Analysis and Manipulation (SCAM)*, Limassol, Cyprus, 116–120.

Breiman, L., Oct. 2001, Random Forests, *Machine Learning*, 1, 45, 5–32.

Cairo, A. S., Carneiro, G. D. F., dan Monteiro, M. P., Nov. 6, 2018, The Impact of Code Smells on Software Bugs: A Systematic Literature Review, *Information*, 11, 9, 273.

Chawla, N. V., Bowyer, K. W., Hall, L. O., dan Kegelmeyer, W. P., June 1, 2002, SMOTE: Synthetic Minority Over-sampling Technique, *Journal of Artificial Intelligence Research*, 16, 321–357.

UNIVERSITAS
GADJAH MADA
**Deteksi Code Smell Menggunakan Kombinasi Metrik Struktural dan Representasi Semantik Kode pada
Dataset yang Relevan dengan Industri**
Azhar Bagaskara, Arif Nurwidyantoro, M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Chen, T. dan Guestrin, C., Aug. 13, 2016, XGBoost: A Scalable Tree Boosting System, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

Chicco, D. dan Jurman, G., Dec. 2020, The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation, *BMC Genomics*, 1, 21, 6.

Feng, Z., Guo, D., Tang, D., Duan, N., Feng, X., Gong, M., Shou, L., Qin, B., Liu, T., Jiang, D., dan Zhou, M., *CodeBERT: A Pre-Trained Model for Programming and Natural Languages*, Sept. 18, 2020, DOI: 10.48550/arXiv.2002.08155, arXiv: 2002.08155[cs], *http://arxiv.org/abs/2002.08155* (diakses pada 11/30/2025).

Fowler, M., 2018, *Refactoring: Improving the Design of Existing Code*, Pearson Education.

Kanade, A., Maniatis, P., Balakrishnan, G., dan Shi, K., July 13, 2020, Learning and Evaluating Contextual Embedding of Source Code, *Proceedings of the 37th International Conference on Machine Learning*, 5110–5121.

Kecman, V., Apr. 22, 2005, *Support Vector Machines – An Introduction*, ed. by Wang, L., red. by Kacprzyk, J., *Support Vector Machines: Theory and Applications*, 177, Springer Berlin Heidelberg, Berlin, Heidelberg.

Khder, M., Dec. 30, 2021, Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application, *International Journal of Advances in Soft Computing and its Applications*, 3, 13, 145–168.

Kirk, D., Roper, M., dan Wood, M., 2007, A heuristic-based approach to code-smell detection, *1st workshop on refactoring tools (WRT'07) proceedings*, United States, 54–56.

Kovačević, A., Slivka, J., Vidaković, D., Grujić, K.-G., Luburić, N., Prokić, S., dan Sladić, G., Oct. 2022, Automatic detection of Long Method and God Class code smells through neural source code embeddings, *Expert Systems with Applications*, 204, 117607.

Lacerda, G., Petrillo, F., Pimenta, M., dan Guéhéneuc, Y. G., Sept. 2020, Code smells and refactoring: A tertiary systematic review of challenges and observations, *Journal of Systems and Software*, 167, 110610.

UNIVERSITAS
GADJAH MADA
**Deteksi Code Smell Menggunakan Kombinasi Metrik Struktural dan Representasi Semantik Kode pada
Dataset yang Relevan dengan Industri**
Azhar Bagaskara, Arif Nurwidyantoro, M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Lewowski, T. dan Madeyski, L., *Creating Evolving Project Data Sets in Software Engineering*, *Integrating Research and Practice in Software Engineering*, ed. by Jarzabek, S., Poniszewska-Marańda, A., dan Madeyski, L., Springer International Publishing, Cham, 1–14, ISBN: 978-3-030-26574-8, DOI: 10.1007/978-3-030-26574-8_1, *https://doi.org/10.1007/978-3-030-26574-8_1*.

Luburić, N., Prokić, S., Grujić, K.-G., Slivka, J., Kovačević, A., Sladić, G., dan Vidaković, D., *Towards a systematic approach to manual annotation of code smells*, Nov. 29, 2021, DOI: 10.36227/techrxiv.14159183.v2, *https://www.techrxiv.org/doi/full/10.36227/techrxiv.14159183.v2* (diakses pada 12/10/2025).

M, H. dan M.N, S., Mar. 31, 2015, A Review on Evaluation Metrics for Data Classification Evaluations, *International Journal of Data Mining & Knowledge Management Process*, 2, 5, 01–11.

Madeyski, L. dan Lewowski, T., Mar. 2023, Detecting code smells using industry-relevant data, *Information and Software Technology*, 155, 107112.

Madeyski, L. dan Lewowski, T., Apr. 15, 2020, MLCQ: Industry-Relevant Code Smell Data Set, *Proceedings of the Evaluation and Assessment in Software Engineering*, Trondheim Norway, 342–347.

Mahmud Sujon, K., Binti Hassan, R., Tusnia Towshi, Z., Othman, M. A., Abdus Samad, M., dan Choi, K., 2024, When to Use Standardization and Normalization: Empirical Evidence From Machine Learning Models and XAI, *IEEE Access*, 12, 135300–135314.

Manning, C. D., Raghavan, P., dan Schütze, H., 2008, *Introduction to information retrieval*, Cambridge university press, Cambridge.

Mohammed, S., Budach, L., Feuerpfeil, M., Ihde, N., Nathansen, A., Noack, N., Patzlaff, H., Naumann, F., dan Harmouch, H., July 2025, The effects of data quality on machine learning performance on tabular data, *Information Systems*, 132, 102549.

Núñez-Varela, A., Perez-Gonzalez, H. G., Cuevas-Tello, J. C., dan Soubervielle-Montalvo, C., 2013, A Methodology for Obtaining Universal Software Code Metrics, *Procedia Technology*, 7, 336–343.

UNIVERSITAS
GADJAH MADA
**Deteksi Code Smell Menggunakan Kombinasi Metrik Struktural dan Representasi Semantik Kode pada
Dataset yang Relevan dengan Industri**
Azhar Bagaskara, Arif Nurwidyantoro, M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Nuñez-Varela, A. S., Pérez-Gonzalez, H. G., Martínez-Perez, F. E., dan Soubervielle-Montalvo, C., June 2017, Source code metrics: A systematic mapping study, *Journal of Systems and Software*, 128, 164–197.

Ogheneovo, E., July 29, 2013, Software Maintenance and Evolution: The Implication for Software Development, *West African Journal of Industrial and Academic Research*, 1, 7, 81–92.

Paiva, T., Damasceno, A., Figueiredo, E., dan Sant'Anna, C., Dec. 2017, On the evaluation of code smells and detection tools, *Journal of Software Engineering Research and Development*, 1, 5, 7.

Palomba, F., Bavota, G., Di Penta, M., Fasano, F., Oliveto, R., dan De Lucia, A., May 27, 2018, On the diffuseness and the impact on maintainability of code smells: a large scale empirical investigation, *Proceedings of the 40th International Conference on Software Engineering*, Gothenburg Sweden, 482–482.

Pecorelli, F., Di Nucci, D., De Roover, C., dan De Lucia, A., Nov. 2020, A large empirical assessment of the role of data balancing in machine-learning-based code smell detection, *Journal of Systems and Software*, 169, 110693.

Pecorelli, F., Di Nucci, D., De Roover, C., dan De Lucia, A., Aug. 27, 2019, On the role of data balancing for machine learning-based code smell detection, *Proceedings of the 3rd ACM SIGSOFT International Workshop on Machine Learning Techniques for Software Quality Evaluation*, Tallinn Estonia, 19–24.

Reis, J. P. d., Abreu, F. B. e., Carneiro, G. d. F., dan Anslow, C., Jan. 2022, Code smells detection and visualization: A systematic literature review, *Archives of Computational Methods in Engineering*, 1, 29, 47–94.

Sharma, T. dan Spinellis, D., Apr. 2018, A survey on software smells, *Journal of Systems and Software*, 138, 158–173.

Škipina, M., Slivka, J., Luburić, N., dan Kovačević, A., June 2024, Automatic detection of Feature Envy and Data Class code smells using machine learning, *Expert Systems with Applications*, 243, 122855.

Sui, Y., Cheng, X., Zhang, G., dan Wang, H., Nov. 13, 2020, Flow2Vec: value-flow-based precise code embedding, *Proceedings of the ACM on Programming Languages*, 4, 1–27.

UNIVERSITAS
GADJAH MADA
**Deteksi Code Smell Menggunakan Kombinasi Metrik Struktural dan Representasi Semantik Kode pada
Dataset yang Relevan dengan Industri**
Azhar Bagaskara, Arif Nurwidyantoro, M.Cs., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Tornhill, A. dan Borg, M., *Code Red: The Business Impact of Code Quality – A Quantitative Study of 39 Proprietary Production Codebases*, Version Number: 1, 2022, DOI: 10.48550/ARXIV.2203.04374, *https://arxiv.org/abs/2203.04374* (diakses pada 11/10/2025).

Zhang, D., Song, S., Zhang, Y., Liu, H., dan Shen, G., Jan. 2024, Code Smell Detection Research Based on Pre-training and Stacking Models, *IEEE Latin America Transactions*, 1, 22, 22–30.

Zhang, Y., Ge, C., Liu, H., dan Zheng, K., Jan. 2024, Code smell detection based on supervised learning models: A survey, *Neurocomputing*, 565, 127014.