UNIVERSITAS
GADJAH MADA
Rancang Bangun Pemindai Hardcoded Secrets Berbasis Shannon Entropy dan Random Forest untuk Mengurangi False Positive
Heaven Aulianisa Pambudi Putri, Dr. Ir. Ronald Adrian S.T., M.Eng., IPM.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR PUSTAKA

[1] S. M. Sutra and A. Haryanto, "Upaya Peningkatan Keamanan Siber Indonesia oleh Badan Siber dan Sandi Negara (BSSN) Tahun 2017-2020," *Global Political Studies Journal*, vol. 7, no. 1, pp. 56–69, Apr. 2023, doi: 10.34010/gpsjournal.v7i1.

[2] J. A. Curtis and N. U. Eisty, "The Role of AI in Modern Penetration Testing," arXiv preprint arXiv:2512.12326, pp. 1-3 Dec. 2025, doi: 10.48550/arXiv.2512.12326.

[3] G. Sánchez, O. Olayinka, and A. Pasikhani, "Web application penetration testing with artificial intelligence: a systematic review," *in 2024 22nd International Symposium on Network Computing and Applications (NCA)*, Bertinoro, Italy, Oct. 2024, pp. 236–245, doi: 10.1109/NCA61908.2024.00043.

[4] "Pembuatan Sistem Deteksi Harcode Credential Pada Repository," ROUTERS: *Jurnal Sistem dan Teknologi Informasi*, vol. 3, no. 1, pp. 20–31, Feb. 2025, doi: 10.25181/rt.v3i1.3898.

[5] M. R. Rahman, N. Imtiaz, M. A. Storey, and L. Williams, "Why secret detection tools are not enough: It's not just about false positives - An industrial case study," *Empirical Software Engineering*, vol. 27, no. 3, p. 59, May 2022, doi: 10.1007/s10664-021-10109-y.

[6] A. Saha, T. Denning, V. Srikumar and S. K. Kasera, "Secrets in Source Code: Reducing *False* Positives using Machine Learning," *2020 International Conference on COMmunication Systems & NETworkS (COMSNETS)*, Bengaluru, India, 2020, pp. 168-175, doi: 10.1109/COMSNETS48256.2020.9027350.

[7] S. K. Basak, L. Neil, B. Reaves, and L. Williams, "SecretBench: A *Dataset* of Software Secrets," *in Proceedings of the 2023 IEEE/ACM 20th International Conference on Mining Software Repositories (MSR),* Melbourne, VIC, Australia, 2023, pp. 312-316, doi: 10.1109/MSR59073.2023.00053.

[8] S. K. Basak, J. Cox, B. Reaves, and L. Williams, "A Comparative Study of Software Secrets Reporting by Secret Detection Tools," *in International Symposium on Empirical Software Engineering and Measurement, IEEE Computer Society*, 2023. doi: 10.1109/ESEM56168.2023.10304853.

[9] R. Muhammad, "*Klasifikasi file berisi secret key menggunakan character level CNN dan Shannon entropy*," M.S. *thesis*, Program Studi S2 Ilmu Komputer, Departemen

UNIVERSITAS
GADJAH MADA
Rancang Bangun Pemindai Hardcoded Secrets Berbasis Shannon Entropy dan Random Forest untuk Mengurangi False Positive
Heaven Aulianisa Pambudi Putri, Dr. Ir. Ronald Adrian S.T., M.Eng., IPM.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Ilmu Komputer dan Elektronika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Gadjah Mada, Yogyakarta, Indonesia, 2023.

[10] R. Han et al., "A Credential Usage Study: Flow-Aware Leakage Detection in Open-Source Projects," *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 722–734, 2024, doi: 10.1109/TIFS.2023.3326985.

[11] W. Wang et al., "An Ensemble Machine Learning Framework for Automated Sensitive Information Leakage Detection in Code Repositories," *Institute of Electrical and Electronics Engineers (IEEE)*, Nov. 2025, pp. 552–556. doi: 10.1109/eit67313.2025.11232238.

[12] M. N. Rahman, S. Ahmed, Z. Wahab, S. M. Sohan, and R. Shahriyar, "Secret Breach Detection in Source Code with Large Language Models," *Institute of Electrical and Electronics Engineers (IEEE)*, Jan. 2026, pp. 207–217. doi: 10.1109/esem64174.2025.00021.

[13] S. Nurul, S. Anggrainy, and S. Aprelyani, "Faktor-Faktor yang Mempengaruhi Keamanan Sistem Informasi: Keamanan Informasi, Teknologi Informasi dan Network (Literature Review SIM)," *Jurnal Ekonomi Manajemen Sistem Informasi*, vol. 3, no. 5, p. 564, May 2022, doi: 10.31933/jemsi.v3i5.

[14] I. Kamilah, Ritzkal, and A. H. Hendrawan, "Analisis Keamanan Vulnerability pada Server Absensi Kehadiran Laboratorium di Program Studi Teknik Informatika," *in Seminar Nasional Sains dan Teknologi (Semnastek)* 2019, Jakarta, Indonesia, Oct. 16, 2019, pp. 1-10.

[15] K. Božić, N. Penevski, S. Adamović, "Penetration Testing and Vulnerability Assessment: Introduction, Phases, Tools and Methods," *in Sinteza 2019 - International Scientific Conference on Information Technology and Data Related Research*, Belgrade, Singidunum University, Serbia, 2019, pp. 229-234. doi:10.15308/Sinteza-2019-229-234

[16] F. Fachri, "Optimasi Keamanan Web Server terhadap Serangan Brute-Force menggunakan Penetration Testing," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK),* vol. 10, no. 1, pp. 51-58, Feb. 2023, doi: 10.25126/jtiik.2023105872.

[17] The MITRE Corporation, "CWE-798: Use of Hard-coded Credentials," Common Weakness Enumeration. [Online]. Available: https://cwe.mitre.org/data/definitions/798.html. (Accessed: Mar. 10, 2026).

UNIVERSITAS
GADJAH MADA
Rancang Bangun Pemindai Hardcoded Secrets Berbasis Shannon Entropy dan Random Forest untuk Mengurangi False Positive
Heaven Aulianisa Pambudi Putri, Dr. Ir. Ronald Adrian S.T., M.Eng., IPM.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

[18] B. R. Chandavarkar, "Hardcoded Credentials and Insecure Data Transfer in IoT: National and International Status," *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, 2020, pp. 1-7, doi: 10.1109/ICCCNT49239.2020.9225520.

[19] H. Harvey, "Behind GitHub's new authentication token formats," *The GitHub Blog*, Apr. 05, 2021. [Online]. Available: https://github.blog/2021-04-05-behind-githubs-new-authentication-token-formats. (Accessed: Mar. 10, 2026)

[20] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379-423, 1948. (Reprinted with corrections 2001)

[21] T. M. Mitchell, *Machine learning*. New York, NY, USA: McGraw-Hill, 1997

[22] A. L. Samuel, "Some Studies in *Machine learning* Using the Game of Checkers," in *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210-229, July 1959, doi: 10.1147/rd.33.0210.

[23] F. Y. Osisanwo, J. E. T. Akinsola, O. Awodele, J. O. Hinmikaiye, O. Olakanmi, and J. Akinjobi, "Supervised Machine learning Algorithms: Classification and Comparison," *International Journal of Computer Trends and Technology, vol. 48, no. 3, 2017.*

[24] L. Breiman, "Random Forests," *Machine learning,* vol. 45, no. 1, pp. 5-32, Oct. 2001, doi: 10.1023/A:1010933404324.

[25] H. A. Salman, A. Kalakech, dan A. Steiti, "Random Forest Algorithm Overview," *Babylonian Journal of Machine learning*, vol. 2024, pp. 69–79, Jun. 2024, doi: 10.58496/BJML/2024/007.

[26] M. Jena and S. Dehuri, "Decision Tree for Classification and Regression: A State-of-the Art Review," *Informatica*, vol. 44, no. 4, pp. 405–420, 2020, doi: 10.31449/inf.v44i4.3023.

[27] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, "Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications," *Journal of Big Data*, vol. 11, no. 113, 2024, doi: 10.1186/s40537-024-00973-y.

UNIVERSITAS
GADJAH MADA
Rancang Bangun Pemindai Hardcoded Secrets Berbasis Shannon Entropy dan Random Forest untuk Mengurangi False Positive
Heaven Aulianisa Pambudi Putri, Dr. Ir. Ronald Adrian S.T., M.Eng., IPM.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

[28] W. Wang, Y. Duan, L. Cao, and Z. Jiang, "Application of improved Naive Bayes classification algorithm in 5G signaling analysis," *The Journal of Supercomputing*, vol. 79, pp. 6941–6964, 2023, doi: 10.1007/s11227-022-04946-x.

[29] N. Amaya-Tejera, M. Gamarra, J. I. Velez, and E. Zurek, "A distance-based kernel for classification via Support Vector Machines," *Frontiers in Artificial Intelligence*, vol. 7, Art. no. 1287875, 2024, doi: 10.3389/frai.2024.1287875.

[30] S. Sathyanarayanan, "Confusion Matrix-Based Performance Evaluation Metrics," *African Journal of Biomedical Research*, pp. 4023–4031, Nov. 2024, doi: 10.53555/ajbr.v27i4s.4345.

[31] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics*, vol. 17, no. 1, pp. 168–192, 2018, doi: 10.1016/j.aci.2018.08.003.

[32] A. Vanacore, M. S. Pellegrino, and A. Ciardiello, "Fair evaluation of classifier predictive performance based on binary confusion matrix," *Comput Stat*, vol. 39, no. 1, pp. 363–383, Feb. 2024, doi: 10.1007/s00180-022-01301-9.

[33] M. M. Fahmy, "Confusion Matrix in Binary Classification Problems: A Step-by-Step Tutorial," *ERJ*, 2022.

[34] K. M. Sujon, R. Hassan, K. Choi, and M. A. Samad, "Accuracy, precision, recall, f1-score, or MCC? empirical evidence from advanced statistics, ML, and XAI for evaluating business predictive models," *J Big Data*, vol. 12, no. 1, Dec. 2025, doi: 10.1186/s40537-025-01313-4.

[35] National Cybersecurity Center of Excellence (NCCoE), "Securing Picture Archiving and Communication System (PACS): Cybersecurity for the Healthcare Sector, Volume A," *NIST*, 2020. [Online]. Tersedia: https://www.nccoe.nist.gov/publication/1800-26/VolA/index.html. [Diakses: 26 Mei 2026].

[36] A. Rashad, "Learning the Basics of Penetration Testing," *Hashnode*, [Online]. Tersedia: https://amrrashad.hashnode.dev/learning-the-basics-of-penetration-testing. [Diakses: 26 Mei 2026].

[37] EnzaTech, "Jenis-Jenis Model Algoritma dalam Machine Learning," *EnzaTech Blog*, Nov. 2023. [Online]. Tersedia: https://enzatech.blogspot.com/2023/11/jenis-jenis-model-algoritma-dalam-machine-learning.html. [Diakses: 26 Mei 2026].

UNIVERSITAS
GADJAH MADA
Rancang Bangun Pemindai Hardcoded Secrets Berbasis Shannon Entropy dan Random Forest untuk Mengurangi False Positive
Heaven Aulianisa Pambudi Putri, Dr. Ir. Ronald Adrian S.T., M.Eng., IPM.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

[38] V. Jain, "Understanding Decision Trees," *Medium*, [Online]. Tersedia: https://medium.com/@jainvidip/understanding-decision-trees-1ba0ef5f6bb4. [Diakses: 26 Mei 2026].

[39] Machine Learning 101, "K-Nearest Neighbors Classifier," *Medium*, [Online]. Tersedia: https://medium.com/machine-learning-101/k-nearest-neighbors-classifier-1c1ff404d265. [Diakses: 26 Mei 2026].

[40] MLArchive, "The Ultimate Guide to Naive Bayes," *MLArchive*, [Online]. Tersedia: https://mlarchive.com/machine-learning/the-ultimate-guide-to-naive-bayes/. [Diakses: 26 Mei 2026].

[41] Analytics Vidhya, "Support Vector Machines (SVM): A Complete Guide for Beginners," *Analytics Vidhya*, Okt. 2021. [Online]. Tersedia: https://www.analyticsvidhya.com/blog/2021/10/support-vector-machinessvm-a-complete-guide-for-beginners/. [Diakses: 26 Mei 2026].

[42] L. Törnqvist, P. Vartia, and Y. O. Vartia, “How should relative changes be measured?,” *The American Statistician*, vol. 39, no. 1, pp. 43–46, 1985, doi: 10.1080/00031305.1985.10479385.

[43] A. J. Vickers, “The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study,” *BMC Medical Research Methodology*, vol. 1, no. 6, 2001, doi: 10.1186/1471-2288-1-6.

[44] J. Mo, H. Kuang, and X. Li, “Password Strength Detection via Machine Learning: Analysis, Modeling, and Evaluation,” *arXiv preprint arXiv:2505.16439*, 2025, doi: 10.48550/arXiv.2505.16439.

[45] Samsung Research, "CredData: A Dataset for Credential Leakage Detection," *GitHub*, 2022. [Online]. Tersedia: https://github.com/Samsung/CredData. [Diakses: 12 April 2026].

[46] K. Thompson, “Regular expression search algorithm,” *Communications of the ACM*, vol. 11, no. 6, pp. 419–422, Jun. 1968, doi: 10.1145/363347.363387.