

REFERENCES

- American Diabetes Association, 2019. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes-2019. *Diabetes Care*, 42:S13–28. <https://doi.org/10.2337/dc19-S002>.
- Afzali, S., & Yildiz, O., 2018. An Effective Sample Preparation Method for Diabetes Prediction. *International Arab Journal of Information Technology*, 15, 968-973.
- Alirezaei, M., Niaki, S. T. A., & Niaki, S. A. A., 2019. A bi-objective hybrid optimization algorithm to reduce noise and data dimension in diabetes diagnosis using support vector machines. *Expert Systems with Applications*, 127, 47–57. doi:10.1016/j.eswa.2019.02.037.
- Arvai, K. 2020. kneed (Version 0.7.0) [Computer software]. <https://doi.org/10.5281/zenodo.6496267>
- Caballé, N., Castillo-Sequera, J., Gomez-Pulido, J. A., Gómez, J., & Polo-Luque, M., 2020. Machine Learning Applied to Diagnosis of Human Diseases: A Systematic Review. *Applied Sciences*, 10, 5135. doi:10.3390/app10155135
- Chakraborty, S., 2011. Analysis and Study of Incremental DBSCAN Clustering Algorithm. *International Journal of Enterprise Computing and Business Systems*, 1.
- Chauhan, T., Rawat, S., Malik, S., & Singh, P., 2021. Supervised and Unsupervised Machine Learning-based Review on Diabetes Care. 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 1, 581–585. doi:10.1109/ICACCS51430.2021.9442021
- Chen, W., Chen, S., Zhang, H., & Wu, T., 2017. A hybrid prediction model for type 2 diabetes using K-means and decision tree. 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), 386–390. doi:10.1109/ICSESS.2017.8342938
- Collier A, Patrick AW, Bell D, Matthews DM, MacIntyre CC, Ewing DJ, Clarke BF., 1989. Relationship of skin thickness to duration of diabetes, glycemic control, and diabetic complications in male IDDM patients. *Diabetes Care*. doi: 10.2337/di-acare.12.5.309. PMID: 2721339.

- D., D., & Sasidhar Babu, D., 2016). Methods to detect different types of outliers. 23–28. doi:10.1109/SAPIENCE.2016.7684114
- Eberle, C., Stichling, S., 2021. Impact of COVID-19 lockdown on glycemic control in patients with type 1 and type 2 diabetes mellitus: a systematic review. *Diabetol Metab Syndr* 13, 95. <https://doi.org/10.1186/s13098-021-00705-9>
- Farajollahi, B., Mehmannaavaz, M., Mehrjoo, H., Moghbeli, F., & Sayadi, M., 2021. Diabetes Diagnosis Using Machine Learning. *Frontiers in Health Informatics*, 10, 65. doi:10.30699/fhi.v10i1.267
- Gayoso, M., Lim, W. Y., Mulekar, M. S., & Kaulfers, A. D., 2021. Effect of Covid-19 quarantine on diabetes Care in Children. *Clinical diabetes and endocrinology*, 7(1), 9. <https://doi.org/10.1186/s40842-021-00122-x>
- Harris, C.R., Millman, K.J., van der Walt, S.J. et al. 2020. Array programming with NumPy. *Nature* 585, 357–362. doi: 10.1038/s41586-020-2649-2.
- Hunter, J. D. 2007. Matplotlib: A 2D graphics environment. *Computing in Science Engineering*, 9(3), 90–95. doi:10.1109/MCSE.2007.55
- Ijaz, M. F., Alfian, G., Syafrudin, M., & Rhee, J., 2018. Hybrid Prediction Model for Type 2 Diabetes and Hypertension Using DBSCAN-Based Outlier Detection, Synthetic Minority Over Sampling Technique (SMOTE), and Random Forest. *Applied Sciences*. doi:10.3390/app8081325
- Ismail, L., Materwala, H., & Al Kaabi, J., 2021. Association of risk factors with type 2 diabetes: A systematic review. *Computational and Structural Biotechnology Journal*, 19, 1759–1785. doi:10.1016/j.csbj.2021.03.003
- Kadhm, M., Ghindawi, I., & Enteesha, D., 2018. An Accurate Diabetes Prediction System Based on K-means Clustering and Proposed Classification Approach. *International Journal of Applied Engineering Research*, 13.
- Khan, K., Rehman, S. U., Aziz, K., Fong, S., & Sarasvady, S., 2014. DBSCAN: Past, present and future. *The Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2014)*, 232–238. doi:10.1109/ICADIWT.2014.6814687

- M. A. Sarwar, N. Kamal, W. Hamid & M. A. Shah, "Prediction of Diabetes Using Machine Learning Algorithms in Healthcare," 2018 24th International Conference on Automation and Computing (ICAC), 2018, pp. 1-6, doi: 10.23919/ICAC.2018.8748992.
- McKinney, W. 2010. Data Structures for Statistical Computing in Python . In Proceedings of the 9th Python in Science Conference (pp. 56 - 61).
- Minks, J., 2020. A theoretical framework to guide a study for exploring the impact of established and potential risk factors for type 2 diabetes mellitus. *Applied Nursing Research*, 53, 151267. doi:10.1016/j.apnr.2020.151267
- NirmalaDevi, M., alias Balamurugan, S. A., & Swathi, U. V., 2013. An amalgam KNN to predict diabetes mellitus. 2013 IEEE International Conference ON Emerging Trends in Computing, Communication and Nanotechnology (ICECCN), 691–695. doi:10.1109/ICE-CCN.2013.6528591
- Osborne, J., & Overbay, A., 2004. The Power of Outliers (and Why Researchers Should Always Check for Them). *Pract. Assess. Res. Eval.*, 9.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Raihan, M., Islam, M. T., Farzana, F., Raju, M. G. M., & Mondal, H. S., 2019. An Empirical Study to Predict Diabetes Mellitus using K-Means and Hierarchical Clustering Techniques. 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 1–6. doi:10.1109/ICCCNT45670.2019.8944552
- Rahmah, N., & Sitanggang, I., 2016. Determination of Optimal Epsilon (Eps) Value on DBSCAN Algorithm to Clustering Data on Peatland Hotspots in Sumatra. *IOP Conference Series: Earth and Environmental Science*, 31, 012012. doi:10.1088/1755-1315/31/1/012012
- Rani, S., Anurag, & Rao, S., 2019. Study and Analysis of Noise Effect on Big Data Analytics. 8, 5841–5850.

- Santhanam, T., & Padmavathi, M. S., 2015. Application of K-Means and Genetic Algorithms for Dimension Reduction by Integrating SVM for Diabetes Diagnosis. *Procedia Computer Science*, 47, 76–83. doi:10.1016/j.procs.2015.03.185
- Salgado C.M., Azevedo C., Proença H., Vieira S.M. (2016) Noise Versus Outliers. In: *Secondary Analysis of Electronic Health Records*. Springer, Cham.
- Sathya, R., & Abraham, A., 2013. Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. *International Journal of Advanced Research in Artificial Intelligence*, 2. doi:10.14569/IJARAI.2013.020206
- Sisodia, D., & Sisodia, D. S., 2018. Prediction of Diabetes using Classification Algorithms. *Procedia Computer Science*, 132, 1578–1585. doi:10.1016/j.procs.2018.05.122
- Sneha, N., Gangil, T., 2019. Analysis of diabetes mellitus for early prediction using optimal features selection. *J Big Data* 6, 13. <https://doi.org/10.1186/s40537-019-0175-6>
- Sujatha, C., 2018. Building Predictive Model For Diabetics Data Using K Means Algorithm.
- Tan, S. Y., Mei Wong, J. L., Sim, Y. J., Wong, S. S., Mohamed Elhassan, S. A., Tan, S. H., ... Candasamy, M., 2019. Type 1 and 2 diabetes mellitus: A review on current treatment approach and gene therapy as potential intervention. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 13(1), 364–372. doi:10.1016/j.dsx.2018.10.008
- Tigga, N., & Garg, S., 2020. Prediction of Type 2 Diabetes using Machine Learning Classification Methods. *Procedia Computer Science*, 167, 706–716. doi:10.1016/j.procs.2020.03.336
- Tran, V., Hirose, O., Saethang, T., Nguyen, L., Dang, X., Le, T., & Satou, K., 2014. D-IMPACT: A Data Preprocessing Algorithm to Improve the Performance of Clustering. *Journal of Software Engineering and Applications*, 07, 639–654. doi:10.4236/jsea.2014.78059
- UCI Machine Learning. 2016. Pima Indians Diabetes Database, Version 1. <https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database>. Accessed 21 December 2021.

- Yılmaz, N., Inan, O., & Uzer, M.. 2014. A New Data Preparation Method Based on Clustering Algorithms for Diagnosis Systems of Heart and Diabetes Diseases. *Journal of medical systems*, 38, 48. doi:10.1007/s10916-014-0048-7
- Zhu, X., & Wu, X., 2004. Class Noise vs. Attribute Noise: A Quantitative Study. *Artificial Intelligence Review*, 22(3), 177–210. doi:10.1007/s10462-004-0751-8