

BIBLIOGRAPHY

- [1] <http://www.depkes.go.id/article/view/18121200001/prevent-prevent-and-prevent-the-voice-of-the-world-fight-diabetes.html>
- [2] InfoDatin (Pusat Data dan Informasi Kementerian Kesehatan Republik Indonesia)
- [3] Vigneswari, D., et al. "Machine Learning Tree Classifiers in Predicting Diabetes Mellitus." 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS). IEEE, 2019.
- [4] Zou, Quan, et al. "Predicting diabetes mellitus with machine learning techniques." *Frontiers in genetics* 9 (2018): 515.
- [5] Mercaldo, Francesco, Vittoria Nardone, and Antonella Santone. "Diabetes mellitus affected patients classification and diagnosis through machine learning techniques." *Procedia computer science* 112 (2017): 2519-2528.
- [6] Tafa, Zhilbert, Nerxhivane Pervetica, and Bertran Karahoda. "An intelligent system for diabetes prediction." 2015 4th Mediterranean Conference on Embedded Computing (MECO). IEEE, 2015.
- [7] Azrar, Amina, et al. "Data mining models comparison for diabetes prediction." *Int J Adv Comput Sci Appl* 9 (2018).
- [8] Saru, S., and S. Subashree. "Analysis and Prediction of Diabetes Using Machine Learning." *International Journal of Emerging Technology and Innovative Engineering* 5.4 (2019).
- [9] Maniruzzaman, Md, et al. "Accurate diabetes risk stratification using machine learning: role of missing value and outliers." *Journal of medical systems* 42.5 (2018): 92.
- [10] Ijaz, Muhammad, et al. "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* 8.8 (2018): 1325.
- [11] Wang, Qian, et al. "DMP_MI: an effective diabetes mellitus classification algorithm on imbalanced data with missing values." *IEEE Access* 7 (2019): 102232-102238.
- [12] Rajni, Amandeep. "RB-bayes algorithm for the prediction of diabetic in "PIMA Indian dataset"." *International Journal of Electrical and Computer Engineering (IJECE)* 9.6 (2019): 4866-4872.
- [13] Shi, Zhan. "Improving k-Nearest Neighbors Algorithm for Imbalanced Data Classification." *IOP Conference Series: Materials Science and Engineering*. Vol. 719. No. 1. IOP Publishing, 2020.

-
- [14] Tyagi, Shivani, and Sangeeta Mittal. "Sampling Approaches for Imbalanced Data Classification Problem in Machine Learning." *Proceedings of ICRIC 2019*. Springer, Cham, 2020. 209-221.
- [15] Devi, R. Delshi Howsalya, Anita Bai, and N. Nagarajan. "A novel hybrid approach for diagnosing diabetes mellitus using farthest first and support vector machine algorithms." *Obesity Medicine* 17 (2020): 100152.
- [16] Nnamoko, Nonso, and Ioannis Korkontzelos. "Efficient treatment of outliers and class imbalance for diabetes prediction." *Artificial Intelligence in Medicine* 104 (2020): 101815.
- [17] Raghavendra, S., and J. Santosh Kumar. "Performance evaluation of random forest with feature selection methods in prediction of diabetes." *International Journal of Electrical & Computer Engineering* (2088-8708) 10 (2020).
- [18] Kumar, N. Komal, et al. "An Optimized Random Forest Classifier for Diabetes Mellitus." *Emerging Technologies in Data Mining and Information Security*. Springer, Singapore, 2019. 765-773.
- [19] Mirza, Shuja, Sonu Mittal, and Majid Zaman. "Decision Support Predictive model for prognosis of diabetes using SMOTE and Decision tree." *International Journal of Applied Engineering Research* 13.11 (2018): 9277-9282.
- [20] Shuja, Mirza, Sonu Mittal, and Majid Zaman. "Effective Prediction of Type II Diabetes Mellitus Using Data Mining Classifiers and SMOTE." *Advances in Computing and Intelligent Systems*. Springer, Singapore, 2020. 195-211.
- [21] Wu, Han, et al. "Type 2 diabetes mellitus prediction model based on data mining." *Informatics in Medicine Unlocked* 10 (2018): 100-107.
- [22] Abd Rahman, Muhammad Hafiz Fazren, Wan Wardatul Amani Wan Salim, and Mohd Firdaus Abd Wahab. "RISK PREDICTION ANALYSIS FOR CLASSIFYING TYPE 2 DIABETES OCCURRENCE USING LOCAL DATASET." *Biological and Natural Resources Engineering Journal* 3.1 (2020): 48-61.
- [23] Moreno-Barea, Francisco J., José M. Jerez, and Leonardo Franco. "Improving classification accuracy using data augmentation on small data sets." *Expert Systems with Applications* 161 (2020): 113696.
- [24] Tanner, Martin A. *Tools for statistical inference: observed data and data augmentation methods*. Vol. 67. Springer Science & Business Media, 2012.
- [25] He, Haibo, and Eduardo A. Garcia. "Learning from imbalanced data." *IEEE Transactions on knowledge and data engineering* 21.9 (2009): 1263-1284.

-
- [26] Madley-Dowd, Paul, et al. "The proportion of missing data should not be used to guide decisions on multiple imputation." *Journal of clinical epidemiology* 110 (2019): 63-73.
 - [27] Kirasich, Kaitlin, Trace Smith, and Bivin Sadler. "Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets." *SMU Data Science Review* 1.3 (2018): 9.
 - [28] Islam, MM Faniqul, et al. "Likelihood prediction of diabetes at early stage using data mining techniques." *Computer Vision and Machine Intelligence in Medical Image Analysis*. Springer, Singapore, 2020. 113-125.
 - [29] Branco, Paula, Luis Torgo, and Rita Ribeiro. "A survey of predictive modelling under imbalanced distributions." *arXiv preprint arXiv:1505.01658* (2015).
 - [30] Brownlee, Jason. *Imbalanced Classification with Python: Better Metrics, Balance Skewed Classes, Cost-Sensitive Learning*. Machine Learning Mastery, 2020.
 - [31] Fernández, Alberto, et al. *Learning from imbalanced data sets*. Berlin: Springer, 2018.
 - [32] Verma, Anurag Kumar, and Saurabh Pal. "Prediction of skin disease with three different feature selection techniques using stacking ensemble method." *Applied Biochemistry and Biotechnology* (2019): 1-20.
 - [33] Kuo, Kuang-Ming, et al. "A multi-class classification model for supporting the diagnosis of type II diabetes mellitus." *PeerJ* 8 (2020): e9920.
 - [34] Khalid, Samina, Tehmina Khalil, and Shamila Nasreen. "A survey of feature selection and feature extraction techniques in machine learning." *2014 Science and Information Conference*. IEEE, 2014.
 - [35] Couronné, Raphael, Philipp Probst, and Anne-Laure Boulesteix. "Random forest versus logistic regression: a large-scale benchmark experiment." *BMC bioinformatics* 19.1 (2018): 270.
 - [36] Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.