|  |  |  |  |
| --- | --- | --- | --- |
| Title | Early-Stage Diabetes Prediction using Data Mining Algorithms | | |
| Author(s) Name | Md Moniruzzaman, A. G. M. Zaman, Rifah Tasnia, Sutopa Biswas, Mehnur Khanam | | |
| Contact Email(s) | agmzaman@aiub.edu | | |
| Published Journal/Conference Name | 2nd International Conference on Computing Advancements - ICCA 2022 | | |
| Type of Publication | Conference | | |
| Volume |  | Issue |  |
| Publisher | ACM Digital Library | | |
| Publication Date | 11 August 2022 | | |
| ISSN | ISBN: 978-1-4503-9734-6 | | |
| DOI | 10.1145/3542954.3542990 | | |
| URL | https://dl.acm.org/doi/10.1145/3542954.3542990 | | |
| Other Related Info. |  | | |
|  | | | |

|  |  |
| --- | --- |
| Abstract |  |
| Diabetes is a very common disease nowadays. If not treated early diabetes can pose a profoundly serious health threat. Much research has been conducted to find out the optimal solution for diabetes detection by applying different data mining algorithms, where the dataset consists of different medicinal attributes. In this study, our aim is to examine whether diabetes can be detected at early-stage by applying different data mining algorithms to the non-medicinal dataset; as well as to investigate whether data normalization techniques can improve the classifiers accuracy.  Naive Bayes, K-Nearest Neighbor (KNN), Support Vector Machines (SVM), Decision Tree, Random Forest, and Gradient Boosting Classifier (GBC) algorithms are applied to the Early Stage Diabetes Risk Prediction Dataset in conjunction with Decimal Point Scaling, Z-Score Normalization, Pareto Scaling, Variable Stability Scaling, Min-Max normalization, Max normalization, Maximum Absolute Scaling, Mean Centered Scaling, Soft-max normalization, Power Transformer, Median and Median Absolute Deviation Normalization, Robust Scaling and Log Scaling normalization methods. In this experiment, we discovered that early-stage diabetes detection is possible without any medical diagnosis data. The result shows that GBC performs better compared to other classification algorithms in combination with data normalization and achieved an impressive 99.038% prediction accuracy. | |