Data Mining Techniques for Dimension Reduction in Healthcare Data

Divya[1*], Prof. Pankaj Saxena[2], Prof. Vineeta Singh[3]

[1]Research Scholar, Department of Statistics, Institute of Social Science,
Dr. Bhimrao Ambedkar University, Agra, Uttar Pradesh, India
[2]Professor, R. B. S. Management Technical Campus, Agra, Uttar Pradesh, India
[3]Professor, Department of Statistics, Institute of Social Science,
Dr. Bhimrao Ambedkar University, Agra, Uttar Pradesh, India

[*]Corresponding Author

ABSTRACT

In the era of the healthcare industry, multiple genetic and environmental factors, acting independently as well as in concert, are considered to be the primary causes of complex diseases. However, as clinical datasets become more diverse, the problem of data redundancy arises. Due to this, the necessary relationships may be hindered. Unnecessary, non-contributed attributes/variables of the data affect the result of the experiment. Here, we used the dimension reduction technique of data mining to reduce the number of irrelevant attributes without jeopardizing the mining result. Datasets may contain many irrelevant features which can be handled by dimension reduction techniques to optimize the results. In this study, different dimension reduction techniques are employed on the UCI Machine Learning Heart disease and Breast Cancer Disease datasets, and the results are analyzed.

Keywords: Data Mining, Dimension Reduction, Heart Disease, Breast Cancer Disease.