

Review of: "Feature Selection and Classification of Type II Diabetes on High Dimensional Dataset"

Shokhan M. Al-Barzinji¹

¹ University of Anbar, Iraq

Potential competing interests: No potential competing interests to declare.

The paper addresses an important issue—classification of diabetes in high-dimensional datasets—and successfully demonstrates the benefits of feature selection. However, improving clarity, expanding on methodology details, and adding a comprehensive analysis of results would significantly enhance the paper's impact. The abstract effectively introduces the context of the study—information mining, dimensionality reduction, and the evaluation of Naïve Bayes classifier performance on Type II Diabetes datasets. The methodology describes the Naïve Bayes classifier and the feature selection techniques applied, such as Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA). The conclusion summarizes that a 4-feature subset significantly improved Naïve Bayes performance, while other subsets with 2 or 6 features yielded mixed results.

However, it could benefit from a clearer structure and the elimination of redundant phrases to improve readability and coherence. Specifically:

Strengths:

1. Explains the importance of feature selection in high-dimensional data analysis.
2. Introduces the use of Naïve Bayes and highlights the dataset (Pima Indian Type II Diabetes).
3. Presents the research focus: comparing classifier performance on different feature subsets.

Weaknesses:

1. Some sentences are overly verbose, which could obscure the main message.
2. The phrase "all of the highlights got during information assortment may not be altogether important to the objective class" is ambiguous. It could be rephrased for clarity.