

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

Spandana Vadloori<sup>1</sup>

<sup>1</sup> National Taipei University of Education

**Potential competing interests:** No potential competing interests to declare.

In the current study, an empirical evaluation is conducted using the Naïve Bayesian classifier on the Pima Indian Type II Diabetes dataset. This evaluation is performed with all features and with a subset of features selected using predefined Python libraries. The performance of the Naïve Bayesian classifier is assessed on these different subsets to understand the impact of high dimensionality on its performance. The author explored this by assessing different subsets to understand the impact of the features. However, this manuscript needs significant improvement in the following areas:

- The manuscript requires corrections for clarity and grammar. The flow of the text is often unclear, making it difficult to follow the author's argument.
- This manuscript currently resembles a homework assignment, with code and results seemingly copied and pasted directly from the screen. I recommend that the author review articles from reputed journals to understand how to present results professionally.
- There is a lack of clear explanation regarding the obtained results. The author should provide a more thorough analysis and interpretation of the findings.
- The author only used the Naive Bayesian classifier. It would be beneficial to include other classifiers in the study and compare their performances. This should provide a more comprehensive understanding of the impact of feature selection.
- The author mentioned that the highest accuracy obtained with 4 features is 63.37%, but Table 1 indicates this result is with 5 features. Additionally, the reported accuracy is low. The author should address this discrepancy and discuss the implications of the low accuracy.