

## **Abstract**

Dealing with data imbalance and outliers is an important challenge in data classification. The aim of this study is to improve classification performance by reducing the effects of class imbalance and the presence of outliers in the dataset. SMOTE-LOF combines the SMOTE oversampling method with the Local Outlier Factor (LOF) to create a synthetic sample that also accounts for potential outliers. Meanwhile, Borderline-SMOTE identifies "borderline" samples in the minority class and then creates synthetic samples along the border between the majority and minority classes. In this study, experiments were conducted using classification algorithms such as Naïve Bayes, and Support Vector Machine on datasets that are imbalanced and contain outliers. The datasets used in this research include Pima Indians, Haberman, Glass, and Rainfall. This research scenario includes a comparison with previous research that has been done regarding SMOTE-LOF and Borderline-SMOTE on the Rainfall dataset. The results showed that on the three datasets, Borderline-SMOTE outperformed SMOTE-LOF on all three classifiers with an average accuracy of 4-6%, precision of 2-4%, recall of 5-10%, and F1 score of 5-6%. When the technique was applied to the Rainfall dataset, the results showed a 10-25% increase in accuracy. The outcomes consistently demonstrate that, when applied to the Pima Indians, Haberman, and Glass datasets, Borderline-SMOTE improves the performance of several classification algorithms. Better accuracy, precision, recall, and F1 score are evidence of this when compared to the application of the SMOTE-LOF technique.

**Kata kunci : Imbalanced data, *smote-lof*, *borderline-smote*, *naïve bayes*, *support vector machine***