

# Under-Sample Binary Data Using CURE for Classification

[T. Kathirvalavakumar](#) , [S. Karthikeyan](#) & [Rajendra Prasath](#)

Conference paper | [First Online: 20 December 2020](#)

203 Accesses

Part of the [Lecture Notes in Computer Science](#) book series (LNAI, volume 11987)

## Abstract

---

Classification is a major break-through in the field of research. The performance of a classifier is highly dependent on the preprocessing. Drawback with most of the classifiers is its performance. It always focuses on the class having a high number of samples and ignores the class having fewer numbers of samples. This problem is identified through state-of-the-art evaluation metrics. To overcome this problem, the data in imbalanced form are converted into balanced form before the classification process. In the proposed work, instead of balancing, samples are re-sampled with the help of cluster based technique CURE. It performs under-sampling by reducing the majority samples but not balancing with minority samples. The experimental results show that the data re-sampled through CURE performs better.

## Keywords

[K-means](#)

[Cure](#)

[Clustering](#)

[Under-sampling](#)

[Class imbalance](#)