Handling Class Imbalance Problem in Software Maintainability Prediction: An Empirical Investigation

Ruchika MALHOTRA, Kusum LATA

Frontiers of Computer Science, DOI: 10.1007/s11704-021-0127-0

Problems & Ideas

Problem:

Problem of prediction of high maintainability effort software classes accurately in the case of imbalanced datasets.

Idea:

To handle the imbalanced data problem for software maintainability prediction (SMP), the data resampling methods, namely Synthetic Minority Oversampling Technique (SM), Spread Subsampling (SS), and Resample with replacement (RR) have been applied and after that effective SMP models are developed with the application of machine learning techniques.

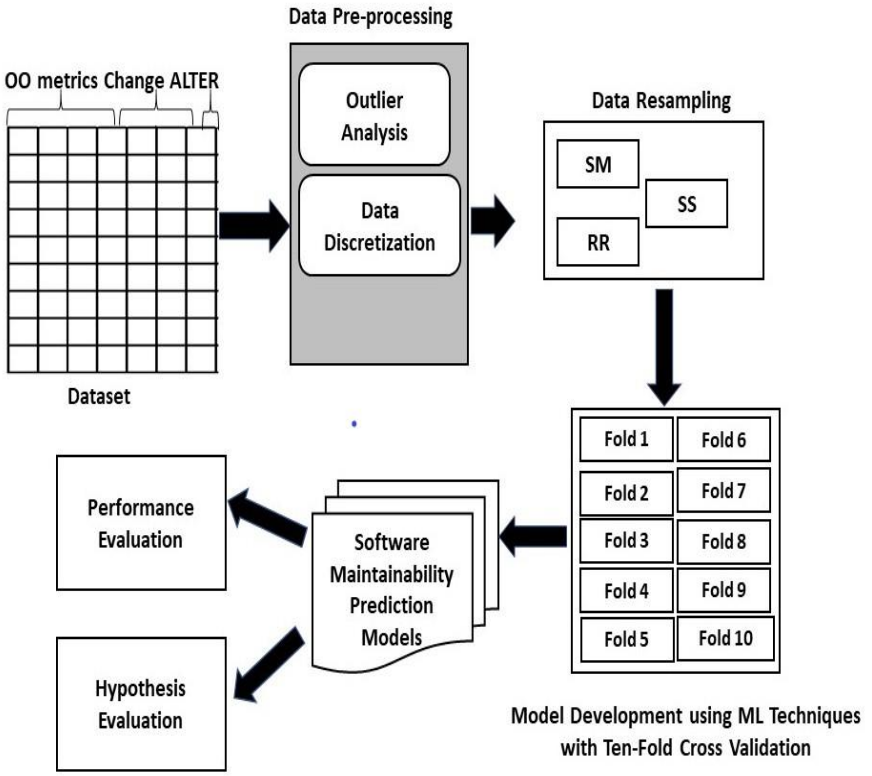

Figure 1: Experimental framework of study

Main Contributions/Conclusions

The major conclusions are summarized as follows:

- The use of data resampling methods, namely RR, SM, and SS, significantly improve the predictions of the SMP models.
- The RR method outperformed other resampling methods, namely SM and SS, for handling imbalanced datasets when the results of the developed models are analyzed with respect to various performance measures.
- Thus, this study supports the RR method to develop efficient SMP models when the dataset in hand is imbalanced.

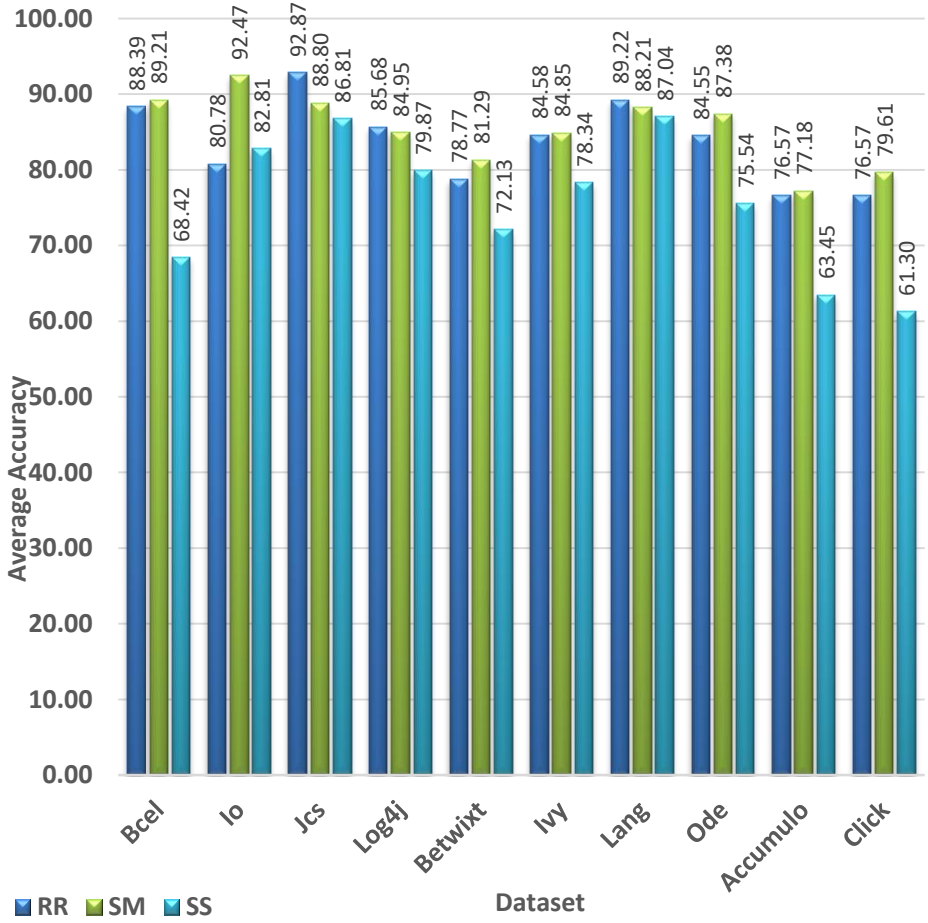

Figure 2 Average Accuracy of SMP mode after data resampling