

EVALUATION OF CLASSIFICATION ALGORITHMS WITH SOLUTION TO CLASS IMBALANCE PROBLEM ON ZAKAT DISTRIBUTION DATASET

Nurul Fazira Ismail and Nor Azriani Mohd Nor
*College of Computing, Informatics and Mathematics,
Universiti Teknologi MARA, Perlis Branch
nurulfazira.ismail01@gmail.com and norazriani@uitm.edu.my*

ABSTRACT - Class imbalance is one of the most serious and significant issues in data mining classification studies. In recent years, there has been a lot of interest in the problem of imbalance in a variety of real-world applications, such as fraud detection, medical diagnosis, and text classification. Similarly, the Zakat distribution UiTM Perlis dataset was utilized for the evaluation class imbalance dataset. The study aims to evaluate the performance of resampling technique, Synthetic Over-Sampling Technique (SMOTE) and to identify the best classifiers for class imbalanced Zakat datasets by comparing the classifiers performance. A sampling-based approach is proposed in this study to solve the imbalance dataset of Zakat distribution in UiTM Perlis. The evaluation is based on various metrics, which include accuracy, precision, recall, F-measure, and the ROC area. The algorithms that are considered in the study include Logistic Regression, Decision Tree (C.4.5), and Random Forest. The results indicate that Random Forest consistently performs well across all evaluation metrics after SMOTE has been implemented. It attains the highest levels of accuracy, precision, recall, F-measure, and ROC area. In conclusion, this research highlights the effectiveness of SMOTE in addressing class imbalance in the distribution of Zakat.

Keywords: Data mining, imbalance dataset, SMOTE, zakat, evaluation classification

1. INTRODUCTION

The zakat distribution dataset at UiTM Perlis shows an imbalance problem exist. The lack of true positive detections has a substantial impact on prediction performance. It implies that the algorithm is not properly capturing or detecting the patterns and properties of positive events. As a result, the algorithm's capacity to anticipate and categorize positive cases in future data may be damaged. Furthermore, it is essential to evaluate the effectiveness of applying the Synthetic Minority Over-Sampling Technique (SMOTE) in addressing the class imbalance problem in zakat distribution. A comparative classify algorithms analysis of sampling methods was conducted. The study outcome on evaluating four classification algorithms for predicting zakat distribution in UiTM Perlis event outcomes.

2. METHODOLOGY

The methodology focuses on addressing the class imbalance problem in the zakat distribution dataset using various techniques. The study utilizes resampling techniques, specifically the Synthetic Minority Over-Sampling Techniques (SMOTE), to improve the balance of class distribution. However, WEKA software was used for the evaluation of various classification algorithms. Weka is a machine learning tool that is written in Java. It provides us with the facility to artificially balance class imbalanced datasets by applying sampling filters in the pre-processing tab (Verma, 2019). The dataset consists of variables such as gender, semester, cumulative grade point average (CGPA), head family's income, mother's income, household size, and zakat receiver. Four classification algorithms, namely logistic regression, decision tree, and random forest are employed to build predictive models. Performance evaluation metrics including accuracy, precision, recall, f-measure, and ROC area are used to assess the model's performance. The results and findings of the evaluation provide insights for selecting the most suitable algorithm for the study.

3. RESULTS AND DISCUSSION

The confusion matrices successfully demonstrated an improvement in detecting the "Not eligible" class, as indicated by the increase in true positive (TP) values for minority class, following the application of SMOTE (Synthetic Minority Over-sampling Technique) to address the class imbalance in the dataset. The updated minority class has 237 members, compared to the actual minority class's 68 members. The instances of the majority class now make up 89% of the dataset, while the minority class only makes up 11%. Based on the supplied data, SMOTE has in this case increased the number of instances in the minority class. This may help to reduce the bias towards the majority class

and increase the representation of the minority class in the dataset. It is possible to draw the conclusion that Random Forest is the best method for the specific scenario when SMOTE is used. It exhibits excellent classification performance as well as successful management of class imbalance. Random Forest consistently demonstrated the highest performance across a variety of evaluation criteria, including accuracy, precision, recall, F-measure, and ROC area. It attains the highest levels of accuracy (95.1181%), precision (0.857), recall (0.676), F-measure (0.756), and ROC area (0.951). Finally, the observed increase in true positive values indicates that SMOTE's implementation has significantly improved the model's ability to detect instances that fall under the "Not eligible" class.

4. NOVELTY OF RESEARCH / PRODUCT

There are many methods available to address the issue of class imbalance in dataset. However, this study introduces an approach by selected various classification algorithms, including Logistic Regression, Decision Tree, and Random Forest, to tackle the class imbalance problem in the specific context of zakat distribution data. Additionally, the study incorporates Synthetic Minority Over-Sampling (SMOTE) to enhance the representation of the minority class and improve classification performance.

5. CONCLUSION

In conclusion, this research highlights the effectiveness of SMOTE in addressing class imbalance in the distribution of zakat. Hence, the results identify Random Forest as the optimal algorithm for the given scenario and provide recommendations for future research to enhance classification performance and investigate additional evaluation techniques.

REFERENCES

- Folorunso, S. O., & Adeyemo, A. B. (2013). Alleviating Classification Problem of Imbalanced Dataset. In *African Journal of Computing & ICT African Journal of Computing & ICT Reference Format S.O. Folorunso & A.B. Adeyemo* (Vol. 6, Issue 2). www.ajocict.net
- Verma, A. (2019). Evaluation of Classification Algorithms with Solutions to Class Imbalance Problem on Bank Marketing Dataset using WEKA. *International Research Journal of Engineering and Technology*. www.irjet.net
- Gupta, S., Saluja, K., Goyal, A., Vajpayee, A., & Tiwari, V. (2022). Comparing the performance of machine learning algorithms using estimated accuracy. *Measurement: Sensors*, 24. <https://doi.org/10.1016/j.measen.2022.100432>