Analyzing Direct Marketing Campaign Using Classification Techniques

M.K.T. Himal[1]*, H.T.K. Abeysundara[2]

[1]*Postgraduate Institute of Science, University of Peradeniya, Peradeniya, Sri Lanka.*
[2]*Department of Statistics & Computer Science, Faculty of Science, University of Peradeniya, Sri Lanka.*
**tharindu.himal9@gmail.com*

Marketing campaigns can be classified into two main types, namely, Mass and Direct. Direct marketing allows banks and other financial organizations to focus on customers who are likely to subscribe to their products, offers, and packages. However, identifying these potential customers often poses a challenge for financial institutions. This study considered a typical bank direct marketing campaign dataset to achieve four main objectives: discuss classification techniques, develop and optimize selected models to predict whether clients subscribe to term deposits or not, determine the key features influencing them, and finally, compare the performances of the developed models to select the best model. Firstly, a comprehensive understanding of classification techniques named Logistic Regression, Decision Tree, Naïve Bayes Approach, Random Forest and K Nearest Neighbor were applied. Secondly, key features were identified that influence a customer's decision to subscribe to a term deposit. Then prediction models were built to predict whether a client will subscribe to a term deposit or not. After the initial model development, data imbalance problem was addressed and the model performances were optimized by tuning the hyper parameters. The significant variables that affect subscription for term deposit were identified as term duration, initiating month, consumer confidence index, number of employees, campaign status, job, consumer price index, euribor-three-month rate, marital status, education, last contact day of the week, outcome of the previous marketing campaign and employment variation rate. Random Forest was identified as the best model which achieved 89% accuracy with 92% sensitivity, 72% specificity with 94% F1 score. Throughout the study, the focus was on maximizing sensitivity while maintaining accuracy, as banks aim to retain all potential customers. The developed model can be used to make suggestions and recommendations for identifying potential customers efficiently and with high accuracy for future scenarios.

Keywords: Classification techniques, Direct marketing campaign, Logistic regression, Sensitivity, Specificity