

Identifying Long-Term Deposit Customers: A Machine Learning Approach

Mohammad Abu Tareq Rony¹, Md. Mehedi Hassan², Eshtiak Ahmed³, Asif Karim^{*4}, Sami Azam⁵, D. S. A. Aashiqur Reza⁶
¹ Department of Statistics, Noakhali Science and Technology University, Noakhali, Bangladesh
² Department of Computer Science and Engineering, North Western University, Khulna, Bangladesh
³ Faculty of Information Technology and Communication Sciences, Tampere University, Tampere, Finland
⁴ College of Engineering, IT and Environment, Charles Darwin University, Darwin, NT, Australia
⁵ Mathematics Discipline, Khulna University, Khulna, Bangladesh
⁶ Emails: abutareqrony@gmail.com¹, mehedi hassan@ieee.org² (ORCID: 0000-0002-9890-0968)², eshtiak.ahmed@tuni.fi³, asif.karim@cdu.edu.au⁴, sami.azam@cdu.edu.au⁵, and aashiq.reza007@gmail.com⁶

Abstract—Majority of the revenue from the banking sector is usually generated from long term deposits by customers. It's very important for banks to understand customer characteristics to increase product sales. To aid this, marketing strategies are employed to target potential customers and let them interact with the banks directly, generating a big amount of data on customer characteristics and demographics. In recent years, it has been discovered using various data analysis, feature selection, and machine learning techniques can be employed to analyze customer characteristics as well as variables that can impact customer decision significantly. These methods can be used to identify consumers in different categories to predict whether a customer would subscribe to a long-term deposit, allowing the marketing strategy to be more successful. In this study, we have taken a R programming approach to analyze financial transaction data to gain insight into how business processes can be improved using data mining techniques to find interesting trends and make more data-driven decisions. We have used statistical analysis like Exploratory Data Analysis (EDA), Principal Component Analysis (PCA), Factor Analysis and Correlations in the given data set. Besides, the study's goal is to use at least three typical classification algorithms among Logistic Regression, Random Forest, Support Vector Machine and K-nearest neighbors, and then make predictive models around customers signing up for long term deposits. Where we have gotten best accuracy from Logistic Regression which is 90.64% as well the sensitivity is 99.05%. Results were analyzed using the accuracy, sensitivity, and specificity score of these algorithms.

Index Terms—Telemarketing, Classification, PCA, Machine learning, Classification