Benha University
Faculty of Commerce
Department of Statistics, Mathematics and Insurance

A Neoteric Prescriptive Statistical Technique on Data Science with Application

By

Noha Nabawy Bahy Ahmed
Assistant Lecturer of Statistics Department of Statistics, Mathematics, And Insurance Faculty of Commerce, Benha University

Supervised by

Prof. Zohdy Mohamed Nofal
Professor and Head Department of Statistics, Mathematics, and Insurance, Faculty of Commerce, Benha University

Dr. Eman Mahmoud Meghawry
Lecturer of Statistics Department of Statistics, Mathematics, and Insurance, Faculty of Commerce, Benha University

Submitted to the Department of Statistics, Mathematics, and Insurance, Faculty of Commerce, Benha University in a partial fulfillment of the requirements for the degree of PHD in Applied Statistics

1445 – 2024

Telecommunications businesses are becoming increasingly competitive in predicting client churn. Because acquiring new customers is expensive, predicting customer attrition is becoming essential, if not indispensable, as it will allow the implementation of a good management policy that will contribute to customer loyalty as well as a strategic planning and decision-making process in the company. This thesis discusses the feasibility of employing various machine learning methods (both coupled and individual) to predict client attrition. The IBM Sample dataset was used for training and testing in this study.

The core goal of this thesis is to employ this strategy to ensure the robustness of the proposed model. For all measurements, no single classifier technique exceeded the others. However, when examining balanced datasets against imbalanced datasets, the combination of various machine learning algorithms and application of the voting technique resulted in measures that exceeded the individual machine learning techniques. F1-score measurements show that the proposed model performs better than the individual models. Furthermore, the results obtained with the IBM imbalanced dataset are statistically significant.

The study proposed model based on a telecommunications sector CCA (customer churn analytics) system. The Statistical algorithm in the proposed architecture handles a feature extraction procedure to locate and implement high-quality mathematical information that aids in lowering the error probability of the feature set. The smaller the error probability training set, the more accurate the trained model. Finding the optimal cost solutions for input feature patterns is an iterative process.

After all the cases have been analyzed, the data element is processed with the prediction model's initialization module. This module takes the optimized data pieces and creates various subsets of the original datasets.

The prediction model is trained and tested using these subsets. In this phase, the training subset and labels for those patterns are processed by the training module. It introduces the combination of clustering, ensemble classification and voting model and stores it in a secondary storage device for loading and evaluation. The test set loads the test subset and the training module before proceeding with the proposed prediction technique.