1. Predicting Customer Churn in the Internet Service Provider (ISP) Industry of Developing Nations: a Single Explanatory Case Study of Trinidad and Tobago.
2. Detecting the Risk of Customer Churn in Telecom Sector: A Comparative Study
3. Customer Churn Prediction with Hybrid Resampling and Ensemble Learning.

|  |  |
| --- | --- |
| Citation Information | 1. Lackeshwar Bachan. University of Liverpool, Liverpool L69 3BX, United Kingdom.  Tarek Gaber. Faculty of Computers & Informatics. Suez Canal University, Ismalia, Egypt. |
| 2. Nabahirwa Edwine, Wenjuan Wang, Wei Song, Denis Ssebuggwawo.  College of Information Technology, Shanghai University, Shangai, China.  Department of Computer Science, Kyambogo University, Kampala, Uganda |
| 3. Takuma Kimura. Hosei University |
| Topicality of Study | 1. Customer churn is rapidly growing issue in the industry. The cost of acquiring new customers focused telecoms sector to ensure that subscriber base retention is optimal. |
| 2. Competition in the telecom industry is increasing and clients can transfer to better and cheaper service. A precise customer churn prediction can help companies to retain their customers. |
| 3. Churn prediction is one of the keys to retain the customers. The technical advancement in data storage and analysis made customer churn one of the most popular topics in the marketing analytics field. |
| Problem | 1. Some works did not produce results that allow for a model to be used in a real-world. Some limitations were the quality and quantity of the dataset. Other work was limited to a singular machine learning model. |
| 2. Some of the results are not satisfactory due to a lack of understanding of datasets and applying inappropriate methods for data analysis. |
| 3. Recent studies showed that ensemble learning methods achieve high performance in classification problems, but only a few studies applied these algorithms in customer churn prediction. In many cases the datasets in prediction is imbalanced. |
| Solution / Aim of Study | 1. Create a model using machine learning that would apply to Trinidad and Tobago industry to predict at-risk customers and their issues. |
| 2. Explore an effective method for detecting the risk of customer churn in telecom sectors through comparing machine learning methods and their optimization algorithms. |
| 3. Develop classification models using novel ensemble algorithms and hybrid resampling and compare them with traditional methods and previous studies. |
| Methodology | 1. Three ML algorithms were applied - Decision Tree, Logistic Regression and Support Vector Machine on a dataset provided by major ISP in Trinidad and Tobago. The performance of these models was evaluated using Accuracy, F1 score and Area under the Curve (AUC). |
| 2. Algorithms of Grid Search, Random Search and Genetic Algorithms were combined to optimize the three machine learning models – Random Forest, Support Vector Machines and K-nearest neighbors. Accuracy, Recall Precision, AUC, F1-score and Mean Absolute Error metrics were used to evaluate performance of the models. |
| 3. The research uses IBM dataset with data of 7043 customers and 21 variables. The training dataset was resampled in three ways: SMOTE, SMOTE Tomek Links and SMOTE-ENN. As classification algorithms this study uses Logistic Regression, SVM, RF, XGBoost, LightGBM and CatBoost. |
| Results and conclusions | 1. Three models were trained and tested. Model Evaluation’s comparative analysis table was described by three indicators – Accuracy, F1-score and AUC.  It concluded that the decision tree model would be the best suited for predicting customer churn. |
| 2. Critical factors related to customer churn as were determined: the length of contract, means of charge and customer service quality. Performance of prediction models was evaluated.  RF algorithm optimized by Grid Search outperformed other models in on both datasets with the maximum accuracy of 99% and 95% on the applied dataset 1-2 respectively. |
| 3. The classification performance was compared using metrics of accuracy, recall, precision, F1-scores and ROC-AUC scores.  In sum, ensemble methods generally outperformed the traditional classification algorithms for all metrics except for SMOTE-ENN resamping in F1-scores. |
| Contribution / Significance of Study | 1. This model would allow the ISP industry to take a proactive approach to address the customer issue. |
| 2. This study can facilitate the construction of an effective prediction tool for managers in the telecom sector and can assist managers to build good relationship with customers. |
| 3. This study showed the superior performance of the proposed models, specifically the combined use of Boosting algorithms with hybrid resampling methods. |