

## Comparative Analysis of Predictive Models for Customer Churn Prediction in the Telecommunication Industry

**Authors :** Deepika Christopher, Garima Anand

**Abstract :** To determine the best model for churn prediction in the telecom industry, this paper compares 11 machine learning algorithms, namely Logistic Regression, Support Vector Machine, Random Forest, Decision Tree, XGBoost, LightGBM, Cat Boost, AdaBoost, Extra Trees, Deep Neural Network, and Hybrid Model (MLPClassifier). It also aims to pinpoint the top three factors that lead to customer churn and conducts customer segmentation to identify vulnerable groups. According to the data, the Logistic Regression model performs the best, with an F1 score of 0.6215, 81.76% accuracy, 68.95% precision, and 56.57% recall. The top three attributes that cause churn are found to be tenure, Internet Service Fiber optic, and Internet Service DSL; conversely, the top three models in this article that perform the best are Logistic Regression, Deep Neural Network, and AdaBoost. The K means algorithm is applied to establish and analyze four different customer clusters. This study has effectively identified customers that are at risk of churn and may be utilized to develop and execute strategies that lower customer attrition.

**Keywords :** attrition, retention, predictive modeling, customer segmentation, telecommunications

**Conference Title :** ICBEF 2024 : International Conference on Business, Economics and Finance

**Conference Location :** Boston, United States

**Conference Dates :** April 22-23, 2024