# Customer churn prediction using machine learning with dimensionality reduction in telecom industry

**Abstract:**

Customer churn is a major problem and one of the most important concerns for large companies. Due to the direct effect on the revenues of the companies, especially in the telecom field, companies are seeking to develop means to predict potential customer to churn. Therefore, finding factors that increase customer churn is important to take necessary actions to reduce this churn. The main contribution of our work is to develop a churn prediction model which assists telecom service providers to predict customers who are most likely subject to churn. The model developed in this work uses machine learning techniques like Logistic Regression, which is a binary classification supervised learning algorithm. Since any dataset considered will be huge in volume in terms of number of data points and also number of attributes or dimensions, we aim to reduce the dimensionality and also to overcome the issues of attribute coreelation on the dataset by Principal Component Analysis before we apply Logistic Regression. We will also analyse the dataset with Exploratory Data Analysis (EDA) techniques and also perform a Model Evaluation to measure the goodness of the model trained.

Key words: Telecom Customer Churn, Logistic Regression, Principal Component Analysis, Machine Learning, Data Science

**Tools/Technologies**:

1. Python 3.x,
2. Python Libraries (numpy, pandas, scikit-learn, seaborn, matplotlib, etc),
3. Anaconda-Navigator,
4. Jupyter Notebook,
5. PyCharm