**PROJECT**

**Customer Churn in Telecommunication Companies**

* **Problem Statement**:

Telecommunication churn refers to customers leaving a telecom service, impacting the company's revenue and profitability. Telecommunications companies face a significant challenge in retaining customers. Customer churn, the loss of subscribers to competing services, not only results in immediate revenue loss but also impacts the long-term profitability and competitiveness of the company. To address this issue, telecommunication companies are actively seeking strategies to identify potential churners and implement targeted retention efforts.

* **Data Collection**:

The data set was provided by ExcelR. Each row corresponds to a client of a company for whom it has collected information about the type of plan they have contracted, the minutes they have talked, or the charge they pay every month.

* **EDA**:

1. Data Cleaning: Data cleaning plays a pivotal role in the Exploratory Data Analysis (EDA) phase, encompassing tasks such as managing missing data and refining the dataset before analysis. By addressing these aspects, you enhance the dependability and precision of insights derived from the EDA. This meticulous process lays a sturdy foundation for subsequent modeling or analysis, heightening their efficiency and robustness. One method frequently employed in this realm is the KNN imputation algorithm, utilized specifically to mitigate null values.
2. Data Visualization: Data visualization in Exploratory Data Analysis (EDA) is a powerful way to understand and explore patterns, relationships, and distributions within a dataset. The plots we use include box plots, scatter plots, bar charts, heatmaps, histogram, pie chart, count plot.

* **Feature Engineering:**

Feature selection is a process used in machine learning to choose the most relevant and informative features (variables) for modeling. In feature selection we can use decision trees, recursive feature elimination(RFE) .

* **Model Building:**

Model building refers to the process of creating and refining a predictive or descriptive model that represents relationships between variables in data. Here we use Decision tree classifier, SVM, Convolutional Neural Networks(CNN), Recurrent Neural Network(RNN), Gradient Boosting Machines(GBM), XGBoost, LightGBM, KNN, Bagging and AdaBoost.

* **Model Evaluation:**

Model evaluation is the process of assessing the performance and effectiveness of a machine learning model. It involves measuring how well the model performs on unseen data or in real-world scenarios.

* **Model Deployment:**

We use STREAMLIT, Flask to make the predictions about the churn.