**Problem Statement: Customer Churn Prediction**

In today's competitive business environment, retaining customers is crucial for sustainable growth and success. This project aims to develop a predictive model to identify customers at risk of churn—discontinuing their use of a service. Customer churn significantly impacts revenue and market share, making it essential to address. By leveraging machine learning techniques, we will create a model that predicts the likelihood of customer churn based on their historical usage patterns, demographic details, and subscription data.

This model will enable the business to proactively engage with high-risk customers using personalized retention strategies, improving customer satisfaction, reducing churn rates, and optimizing business performance. The ultimate goal is to deliver a solution that enhances customer loyalty and fosters long-term success for the organization.

**Data Description**

The dataset provides customer information for predicting churn and contains the following columns:

**CustomerID:** Unique identifier for each customer.

**Name**: Name of the customer.

**Age**: Age of the customer.

**Gender**: Gender of the customer (Male or Female).

**Location**: Customer’s location, with options including Houston, Los Angeles, Miami, Chicago, and New York.

**Subscription\_Length\_Months**: Number of months the customer has been subscribed.

**Monthly\_Bill:** Monthly billing amount.

**Total\_Usage\_GB**: Total data usage in gigabytes.

**Churn**: Binary indicator (1 for churned, 0 for retained).

**Technologies and Techniques Used**

Programming Language

Python: Chosen for its rich ecosystem of libraries and ease of use in data analysis, machine learning, and visualization.

Data Manipulation and Analysis

**Pandas:** Used for data manipulation and processing.

**NumPy:** Provides numerical computing support, particularly for handling multi-dimensional arrays.

Data Visualization

**Matplotlib:** Creates static, interactive, and animated plots.

**Seaborn:** Simplifies creating attractive statistical visualizations.

**Machine Learning Tools**

**Scikit-Learn:** Provides tools for classification, regression, clustering, model evaluation, and selection.

**Random Forest Classifier:** Used as an ensemble learning algorithm to improve model accuracy and robustness.

**XGBoost:** Boosting algorithm designed for speed and performance.

**Variance Inflation Factor (VIF):** Detects multicollinearity among predictors.

**Data Preprocessing**

**StandardScaler:** Standardizes data to improve model performance.

Principal Component Analysis (PCA): Reduces dimensionality while preserving variance.

**Classification Algorithms**

Logistic Regression, Decision Tree, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naive Bayes, AdaBoost, Gradient Boosting, XGBoost, and Neural Networks were explored to find the best-performing model.

**Deep Learning Frameworks**

**TensorFlow and Keras:** Used for building and training neural network models.

**Model Evaluation and Tuning**

Evaluation Metrics: Accuracy, precision, recall, F1-score, confusion matrix, ROC curve, and AUC (Area Under Curve).

**GridSearchCV:** Hyperparameter tuning through exhaustive parameter search.

**Cross-Validation:** Evaluates model generalization across subsets of data.

**Early Stopping:** Prevents overfitting during neural network training.

**ModelCheckpoint:** Saves the best model during training.

**Outcome**

The project results in a machine learning model capable of predicting customer churn. Using features like age, gender, location, subscription length, monthly bill, and total usage, the model identifies customers at high risk of churn. This predictive capability empowers the business to:

**Implement targeted retention strategies.**

Personalize customer engagement.

Allocate resources effectively.

By reducing churn rates and improving customer satisfaction, this solution supports the organization's goal of long-term growth and customer loyalty. Success is measured by the model's predictive accuracy and its ability to contribute to reduced churn rates and optimized business strategies.