Customer Churn Prediction:

Project Summary:

The objective of this project is to build a machine learning model that predicts customer churn based on historical customer data. The project follows a standard machine learning project pipeline, encompassing data preprocessing, feature engineering, model selection, optimization, and deployment. The dataset contains customer attributes, including age, subscription length, monthly bill, total usage, gender, and location.

Flow of the work:

- Data Loading and First look exploration
- Exploratory Data Analysis
- Feature engineering and Scaling
- Variables spliting
- Model Development
- Pickling the required model and dependencies.
- Model Deployment (VS code)

Approach:

1. Data Preprocessing:

- Data Exploration: The dataset was initially loaded and explored to understand its structure and characteristics.

- Data Cleaning: Missing data and outliers were meticulously handled to ensure data quality and minimize the impact of anomalies on model performance.

- Encoding Categorical Variables: Categorical variables (Gender and Location) were encoded. Gender was label encoded, while Location was one-hot encoded.

- Data Splitting: The dataset was divided into training and testing sets to facilitate model development and evaluation.

2. Feature Engineering:

- Relevant Feature Generation: Additional features such as "Amount_Paid" and "Usage_Per_Month" were engineered to potentially enhance the model's predictive power.

- Feature Scaling and Normalization: Numerical features were scaled and normalized where necessary to align with the requirements of specific algorithms.

3. Model Building:

- Model Selection: Three machine learning algorithms were chosen for model development: Logistic Regression, XGBoost, and Support Vector Machine (SVM).

- Model Training and Validation: Each model was trained and validated using the training dataset.

- Model Evaluation: Model performance was assessed using various metrics, including accuracy, precision, recall, and F1-score.

4. Model Optimization:

- Hyperparameter Tuning: Model parameters were fine-tuned to optimize predictive performance.

- Cross-Validation: Techniques like cross-validation were employed to ensure robustness and assess generalization.

5. Model Deployment:

- Production Deployment: The XGBoost model was selected as the final model and deployed in a simulated production environment using Flask.

- Real-time Predictions: Users can input customer data, which is preprocessed to match model features, and receive churn predictions in real-time.

Model Performance:

The XGBoost model achieved the following accuracy metrics:

- Accuracy: 0.55250

- Precision: 0.544647547011241

- Recall: 0.4811623246492986

- F1-score: 0.4925380788758397

Business Understanding and Demand:

It's important to note that while the model has achieved reasonable predictive performance, further business understanding and domain-specific insights can significantly enhance the analysis. Understanding customer behavior, preferences, and external factors that may influence churn is crucial for a more comprehensive analysis. Additionally, feedback from stakeholders and end-users can drive refinements and improvements in the analysis, leading to more accurate and actionable predictions.

In conclusion, this project successfully addressed the challenge of customer churn prediction through a well-structured machine learning pipeline. The choice of the XGBoost model, coupled with feature engineering and preprocessing, yielded reasonable results. However, continuous refinement, business engagement, and a deeper understanding of customer behavior will contribute to the continued success of the churn prediction system.