

CUSTOMER CHURN PREDICTION

Introduction:

In this phase, we will transform the design concept for customer churn prediction into an innovative solution to the problem. Steps involved in transforming design into innovative solution to the problem :

Step 1: Define the Problem

1. Understand Business Objectives: Clearly define the business objectives related to customer churn and how predicting churn can add value.
2. Define Churn: Specify what constitutes churn (e.g., subscription cancellation, non-renewal, inactivity for a certain period).

Step 2: Gather and Understand Data

1. Data Collection: Collect relevant data including customer demographics, transaction history, customer service interactions, etc. Check the various attributes of data like shape (rows and cols), Columns, datatypes
2. Handling missing data:
 - For features with less missing values- can use regression to predict the missing values or fill with the mean of the values present, depending on the feature.
 - For features with very high number of missing values- it is better to drop those columns as they give very less insight on analysis.
 - As there's no thumb rule on what criteria do we delete the columns with high number of missing values, but generally you can delete the columns, if you have more than 30-40% of missing values. But again there's a catch here, for example, Is_Car & Car_Type, People having no cars, will obviously have Car_Type as NaN (null), but that doesn't make this column useless, so decisions has to be taken wisely.
3. Data Exploration: Perform exploratory data analysis (EDA) to understand data distributions, patterns, and correlations.
 - Plot distribution of individual predictors by churn.
 - Perform univariate and bivariate analysis.

4. Data Cleaning:

- Create a copy of base data for manipulation & processing.
- Remove columns not required for processing.

Step 3: Feature Engineering

1. Create Relevant Features: Generate new features that might be indicative of customer behaviour, like customer tenure, average transaction value, or frequency of interactions.

2. Feature Selection: Choose features that have the most impact on churn prediction. Use techniques like correlation analysis or feature importance from models.

Step 4: Data Preprocessing

1. Data Transformation: Standardize or normalize features as necessary, especially for algorithms sensitive to scale.

2. Data Splitting: Divide the data into training and testing sets.

Step 5: Model Selection

1. Choose Algorithms: Select appropriate algorithms (e.g., logistic regression, decision trees, random forests, neural networks) and check which gives more accuracy.

2. Model Training: Train different models using the training data.

3. Pickling the model: Serialize the selected machine learning model and associated objects into a binary format for future use, facilitating reusability, scalability, and efficiency in deployment and predictions.

Step 6: Model Evaluation

1. Metrics Selection: Choose evaluation metrics (e.g., accuracy, precision, recall, F1-score, ROC AUC)

2. Cross-Validation: Use techniques like k-fold cross-validation to assess the models' performance robustly.

Step 7: Hyperparameter Tuning

1. Grid Search or Random Search: Perform hyperparameter tuning using techniques like grid search or random search to optimize the models.
2. Validation Data: Use a separate validation dataset to fine-tune the hyperparameters to prevent overfitting.

Step 8: Model Validation and Testing

1. Validation: Validate the tuned model using the validation dataset to ensure it generalizes well.
2. Testing: Assess the final model using the test dataset, which it has never seen before, to get a real-world performance estimate.

Step 9: Deployment

1. Deployment Environment: Deploy the validated model in the production environment where it can make real-time predictions.
2. Monitoring: Implement a monitoring system to track the model's performance over time and detect drift or degradation.

Step 10: Model Maintenance

Model Maintenance: Regularly update the model using new data to ensure its predictions remain accurate and relevant.