Assignment 1: Customer Churn Prediction for Tele Company

Name : Omar ahmad abdelaziz ahmad

Data Science workflow

1. Problem Definition and Understanding

Objective: Define what constitutes "churn" for your business—whether it's account cancellation, lack of engagement, or inactive status over a defined period.

Goal Setting: Set a clear objective, like reducing churn by a specific percentage in three months.

2. Data Collection

Data Sources: Gather data from CRM, transaction logs, customer service records, engagement metrics, and other relevant sources.

Customer Behavior Data: Look for trends in product usage, frequency of interactions, and previous purchase history, which are often key indicators of churn risk.

3. Data Cleaning and Preprocessing

Data Quality Check: Address missing values, detect and handle outliers, and ensure consistency. Impute missing values where possible, especially for critical features.

Feature Engineering: Create features that reflect potential risk factors, such as recent engagement frequency, unresolved issues, or last interaction time.

Data Transformation: Scale or normalize data as needed, and handle imbalanced classes (e.g., through SMOTE or undersampling) to improve model training on the minority class (churn).

4. Model Selection and Training

Model Choice: Choose algorithms suitable for classification, such as logistic regression, decision trees, random forests, or gradient boosting. These methods often perform well in churn prediction.

Hyperparameter Tuning: Use grid search or random search for optimal parameter settings, improving model performance without overfitting.

5. Model Evaluation

Evaluation Metrics: Use metrics like recall, precision, F1-score, and AUC-ROC to understand how well the model captures at-risk customers while balancing false positives and negatives.

Interpretability: Use SHAP or LIME to interpret feature importance, identifying which factors most influence churn risk.

6. Deployment and Development

Deployment: Integrate the model into a live system (e.g., CRM) to flag at-risk customers, allowing for timely retention actions like targeted marketing or outreach.

Monitoring and Feedback: Track model performance and customer responses, especially on metrics like retention lift or re-engagement rate, to evaluate the impact of the intervention.

Continuous Improvement: Periodically retrain and refine the model with new data and feedback from interventions to maintain its accuracy and relevance as customer behaviors evolve.