PREDICTING TELECOM CUSTOMER CHURN IN INDIAN MARKETS

Presented By Kartik Waghela

Intro

Project Overview

This presentation focuses on predicting telecom customer churn in the Indian and Southeast Asian markets using data analytics techniques. The objective is to retain high-value customers by identifying churn patterns and implementing proactive retention strategies.

Business Objective and the Data

The business objective entails predicting customer churn in the ninth month based on behavior observed in the preceding three months. Understanding customer behavior during churn is critical, characterized by phases including 'good', 'action', and 'churn'. Data Dictionary provides essential meanings of abbreviations used in the dataset.

Data Preparation

- **Step 1**: High-value customers are identified by filtering based on recharge amount, ensuring focus on significant revenue contributors.
- **Step 2**: Churners are tagged based on their activity (or lack thereof) in the churn phase, and attributes related to this phase are removed to prevent bias in predictive modeling.

Modelling Approach

- **Step 3**: Predictive models are built to forecast customer churn, aiding in proactive retention strategies.
- **Techniques**: Class imbalance is addressed to ensure accurate predictions. Techniques like logistic regression is employed for their suitability in handling predictive tasks.

Evaluation and Recommendations

- **Step 4**: Predictive models' performance is assessed to ensure reliability and effectiveness.
- **Step 5**: Important predictor attributes are visualized to understand their impact on churn prediction.
- **Step 6**: Recommendations are provided based on insights gained, enabling telecom companies to implement targeted strategies for managing customer churn effectively.

Important Features Identified

The identified features, including 'good_action_loc_ic_t2m_mou', 'good_action_total_ic_mou', 'good_action_std_ic_mou', 'good_action_roam_og_mou', 'good_action_last_day_rch_amt', and 'rech_data_8', are crucial predictors of potential churn customers as per Lasso and AdaBoost algorithms. Reduction in incoming and outgoing calls usage, along with patterns of recharges, are key indicators during the 'action' phase compared to the 'good' phase.

PCA Results

Principal Component Analysis (PCA) yields a reasonable F1 score of 0.85 on the test data, along with precision of 0.88 and recall of 0.82, indicating satisfactory performance in predicting customer churn.

Lasso Regression Approach

Lasso regression with a mild regularizer (0.01) is employed to identify major predictors without immediately discarding them. Non-important and highly correlated features are preliminarily eliminated through Recursive Feature Elimination (RFE) for optimal parameter selection.

Overfitting Analysis

Overfitting is observed, indicated by a near-perfect F1 score for training data but lower performance on validation data. GridSearch is utilized for parameter optimization to mitigate overfitting issues.

Correlation Analysis

The correlation matrix highlights significant correlations between various features, providing insights into customer behavior and preferences. Key correlations indicate dependencies between recharge amounts, call usage patterns, and data usage preferences.

THANK YOU