LOGISTIC CLASSIFIER FOR PREDICTING CUSTOMER CHURN IN SYRIATEL

Robinson Mumo

INTRODUCTION

- Telecom Industry Challenges: The relentless competition necessitates robust customer retention strategies crucial for sustained growth in the telecommunications sector.
- Companies must leverage on technology to understand their customer needs and anticipate them to fashion tools, products and promotion to meet those needs.

BUSINESS UNDERSTANDING

- **Importance of Customer Retention:** Long-term profitability and brand loyalty hinge on effectively retaining customers amidst market flux and competitive offerings.
- **Strategy:** Leveraging Data Science insights to optimize customer engagement strategies, a key driver in bolstering customer loyalty.

PROBLEM STATEMENT & OBJECTIVES

Challenge:

Precise churn prediction is pivotal for preemptive measures, curbing revenue loss, and reinforcing customer retention initiatives.

Objectives:

Develop a predictive model to determine churn likelihood with a targeted accuracy of 85% based on diverse observations.

Identify and analyze the primary features influencing churn rates within Syriatel's customer base.

DATASET OVERVIEW:

- Features & Target Variable: The dataset comprises numerous customer-centric features observed for churn analysis.
- Source: Kaggle Dataset Link: https://www.kaggle.com/datasets/becksddf/churn-in-telecoms-dataset
- Data Cleaning: The dataset did not have missing values or duplicates
  - Cleaned a few categorical columns to make them numeric (using category encoding)
  - Set phone number as index

EXPLORATORY DATA ANALYSIS (EDA)

- Utilized custom distribution and count plots functions to visualize data distributions.
- Observed that continuous variables mostly followed a normal distribution, showcasing Gaussian curves.
- Identified insights regarding the distribution patterns of different features and their potential impact on the project's objectives.

FEATURE ENGINEERING & ANALYSIS

- We created additional features like "Average Call Duration for Different Call Times," "Customer Interaction Index," and "Total Activity Index" to capture nuanced customer behaviors.
- The newly engineered features provide insights into intricate customer behavior patterns, revealing crucial churn predictors.
- Feature engineering led to the identification of potential churn indicators, guiding focused retention strategies.

DATA PREPROCESSING

Data Splitting

- Split the dataset into 'train_df' and 'test_df' using sklearn for modeling purposes.

Correlation Analysis

- Employed the chi-squared test for categorical features.
- Conducted point-biserial correlation for continuous values.
- Eliminated features with p-values > 0.05.

DATA PREPROCESSING

- **Collinearity Analysis**
- Developed a function to handle collinear features.
- Obtained a 'clean_df' with shape (2499, 10) after dropping certain features due to correlation and collinearity.
- **Class Imbalance Handling**
- Employed SMOTE technique to address class imbalance in the dataset.
- **Dimensionality Reduction (PCA)**
- Utilized PCA (Principal Component Analysis) for dimensionality reduction.
- Determined the number of components for PCA to be 9.

MODELING

- Utilized various machine learning models:
  - **Logistic Regression**
  - **DecisionTreeClassifier**
  - **KNNs**
  - **Random Forests**
  - **XG-Boost Classifier**
  - **Ensembling (Meta-learner based on all the above models)**

MODEL PERFORMANCE

- **XG-Boost**: Achieved an accuracy of 92.45% with notable precision and recall scores, indicating good overall performance.
- **- Random Forest**: Slightly lower accuracy at 91.61% but balanced precision and recall scores.
- **KNN**: Lower accuracy at 82.37% with lower precision and moderate recall.
- **Decision Tree**: Performed less accurately at 76.74% with lower precision and recall.
- **Meta-learner**: Performed the best with an accuracy of 93.17%, high precision, recall, and F1 score.

TOP PREDICTORS OF CHURN RATE

- **International Plan & Voice Mail Plan**: Customers subscribed to an international plan or a voicemail plan may exhibit different usage patterns or behavior leading to a higher likelihood of churn. For instance, dissatisfaction with international call quality or not utilizing voicemail services could influence churn.
- **Total Day Charge, Total Intl Calls, Total Intl Charge**: Higher charges incurred during the day or for international calls might lead to dissatisfaction, prompting customers to consider switching service providers.
- **Customer Service Calls**: Customers frequently reaching out to customer service might indicate issues or dissatisfaction with the service, which could correlate with a higher churn rate.
- **Day Charge-Minute Ratio & Intl Charge-Minute Ratio**: Higher ratios of charges to minutes might signify expensive plans relative to usage, potentially causing dissatisfaction and prompting customers to churn.
- **Resident State**: State information could capture regional tendencies or variations in customer behavior that influence churn rate, like local service quality or competitive offerings in different areas.

CONCLUSION & RECOMMENDATIONS

- **Focus on Customer Experience:** Addressing international calling service concerns and improving customer service interactions directly impacts satisfaction, reducing churn rates.
- **Customized Plans:** Align pricing structures with actual usage to prevent dissatisfaction from high charges, thus reducing the likelihood of churn.
- **Enhance Service Quality:** Address regional disparities in service quality to mitigate churn, considering varying satisfaction levels across different areas.
- **Promote Engagement:** Encourage value-added services like voicemail plans, enhancing customer engagement and potentially reducing churn rates.
- **Continuous Monitoring:** Regularly adapt strategies based on evolving customer needs, ensuring prompt adjustments aligned with customer preferences.

MODEL DEPLOYMENT

- Saving Trained Model: Utilized joblib to save the trained model, ensuring accessibility and ease of deployment.
- Streamlit Application: Developed a 'streamlit.py' file enabling interaction with the pre-trained model, facilitating predictions and analysis.

NEXT STEPS

- Continuous Monitoring: Continue monitoring customer feedback and usage patterns, adapting strategies to evolving needs.
- Strategy Evolution: Continual evolution of strategies for churn mitigation based on ongoing customer behavior analysis.