PRCL-0017- TELECOM CHURN

PRESENTED BY:

- ANUPRIYA S R
- PRATHAM P M
- ANJALI M K

PROJECT TEAM ID - PTID-CDA-NOV-24-246

Model Comparision

By using grid cv and fitting best parameters using hyper-parameter tuning for different models, we got the best score of, KNN: 70.12%, Random Forest: 92.42%, Logistic Regression: 85.93%, Greadient Bosting: 88.74, XG Boost: 92.64%.

- Therefore from the above models **Random Forest** and **XG Boost** perform well with the highest accuracy score of **92.42%** and **92.64%**.

Data Overview:

- The dataset contains information such as account length, service plans (International/VMail), call minutes, customer service calls (as an indicator of dissatisfaction), and churn (the target variable).
- Features like daily, evening, night, and international call durations and charges were present, along with demographic features like state and area code.

Business Case:

- No-Churn Telecom is an established Telecom operator in Europe with more than a decade in Business. Due to new players in the market, telecom industry has become very competitive and retaining customers becoming a challenge.
- In spite of No-Churn initiatives of reducing tariffs and promoting more offers, the churn rate (percentage of customers migrating to competitors) is well above 10%.
- No-Churn wants to explore possibility of Machine Learning to help with following use cases to retain competitive edge in the industry.

PROJECT GOAL :

1. Understanding the variables that are influencing the customers to migrate.

2. Creating Churn risk scores that can be indicative to drive retention campaigns.

3. Introduce new predicting variable “CHURN-FLAG” with values YES(1) or NO(0) so that email campaigns with lucrative offers can be targeted to Churn YES customers.

Data Preprocessing:

- **Data Cleaning**:
    - Handling missing values by filling, imputing, or removing them.
    - Removing duplicates and correcting errors in the data.

- **Data Transformation**:

    - Normalizing or scaling numerical features to bring them into a comparable range.
    - Encoding categorical variables using techniques like one-hot encoding or label encoding.

Feature Engineering:

- Call minutes were converted to hours.
- New features such as total daily calls and charges per hour were created.

Handling Outliers:

- Outliers were detected and replaced with mean values using IQR.

Imbalanced Data:

- The dataset was imbalanced, with more non-churn cases than churn cases.
- SMOTE (Synthetic Minority Oversampling Technique) was used to balance the target variable.

Exploratory Data Analysis:

- Univariate and bivariate analyses were conducted using histograms, scatter plots, and heatmaps.

Key Findings:

- International calls were slightly skewed.
- High charges during the day compared to the evening and night might indicate a potential reason for churn.
- Customer service call frequency strongly correlates with churn.
- Certain states showed higher churn rates.

Modeling and Evaluation:

Several machine learning models were tested, and their performance was evaluated using accuracy, confusion matrix, and ROC-AUC scores.

Models Used:

1. Logistic Regression
2. Decision Tree
3. Random Forest
4. K-Nearest Neighbors (KNN)
5. Support Vector Classifier (SVC)
6. XGBoost

Model Performance:

- ❖ Random Forest: 92.42% accuracy
- ❖ XGBoost: 92.64% accuracy
- ❖ Logistic Regression: 85.93% accuracy
- ❖ Gradient Boosting: 88.74% accuracy
- ❖ KNN: 70.12% accuracy

The **Random Forest** and **XGBoost** models performed the best, achieving accuracies of **92.42%** and **92.64%** respectively. Hyper-parameter tuning using GridSearchCV helped optimize these models for maximum accuracy.

Conclusion:

- ✓ This churn prediction model provides No-Churn Telecom with actionable insights by identifying key churn drivers and providing churn risk scores for each customer.
- ✓ The high accuracy of the Random Forest and XGBoost models will allow the company to focus its retention strategies on customers with the highest likelihood of churning,
- ✓ ultimately improving customer retention and reducing churn rates.