PIDC- HACKATHON

Problem: 1.2 Build a predictive model to estimate the likelihood of customer churn for a telecommunication company. The model should consider factors such as customer demographics, usage patterns, billing history, and customer service interactions.

Library used: numpy, pandas, seaborn, matplotlib, sklearn

Models used: logistic regression, SVM, XGboost Model, Random forest

Telco Customer Churn Prediction is a classification machine learning project aimed at predicting customer churn for a telecommunication company. Customer churn refers to customers who have left the company within the last month, labeled as 'yes' or 'no' in the dataset.

Step 1: Data Collection and Preprocessing
Dataset:
https://drive.google.com/file/d/1dizhBy_bXTwe_muof-oSzndKcbWgpIEY/view?usp=sharing

- Conversion of the 'TotalCharges' column to float type and handling missing values.
- No duplicate data found.

Step 2: Exploratory Data Analysis (EDA)

- Analyzed the distribution of churn rate and found an imbalance in the data.
- Visualized categorical features' count plots and histograms/box plots for continuous features.

Step 3: Feature Encoding
Tested several encoding techniques and found One-Hot encoding to be the most effective.

Step 4: Feature Engineering
- Binned the 'tenure' feature into 6 ranges.
- Applied log transformation to 'MonthlyCharges' and 'TotalCharges' features.

Step 5: Data Imbalance Handling
Used SMOTE (Synthetic Minority Oversampling Technique) to balance class distribution.

Step 6: Model Evaluation
- Four different models were applied on the data and all results are reported with confusion matrix and classification report showing the precision, recall, and f1-score metrics.

- Logistic regression Best parameters after several trials: C=200 (very large c value trying to fit the data as possible without overfitting), max_iter=1000
- Support vector classifier Best prameters: kernel='linear', C=20
- XGBoost classifier RandomizedSearchCV is used for hyperparameters tuning with StratifiedKFold of 5 splits.

Step 7: Model Deployment

- Deploy the trained model in a production environment for real-time predictions.
- Choose a deployment approach based on scalability, resource requirements, and infrastructure preferences:
- Cloud-based deployment using platforms like AWS, Azure, or GCP.
- On-premises deployment within the company's infrastructure.
- Develop APIs or web services to integrate the predictive model with existing systems or applications.

Note: We don't have a source to deploy it on my system please deploy it from your side.

Step 8: Monitoring and Maintenance

- Set up monitoring mechanisms to track model performance, detect drifts in data distribution, and identify any degradation in prediction accuracy.
- Implement regular model retraining cycles to keep the model up-to-date with changing customer behavior and market dynamics.

Conclusion:

In conclusion, the Telco Customer Churn Prediction project successfully developed machine learning models to predict customer churn based on various features such as service subscriptions, account information, and demographic data through data preprocessing, feature engineering, and model training, we were able to achieve promising results in predicting customer churn.

- Logistic Regression achieved an accuracy of approximately 74%.
- Support Vector Classifier achieved an accuracy of approximately 74%.
- XGBoost Classifier achieved an accuracy of approximately 79%.
- Random forest achieved an accuracy of approximately 78%.

    Performed model optimization using GridSearchCV with the Random Forest model Which resulted in increasing model accuracy to 80%.