Telecom Customer Churn Prediction

Develop a predictive model to identify customers likely to leave a telecom provider

Pragathi M Porawakara Arachchige
DSC680-T301 Applied Data Science

Business Problem

- **High churn rates lead to revenue loss and increased customer acquisition costs**
- **Identifying at-risk customers early is crucial for business sustainability**
- **Need for a data-driven approach to improve retention decision-making**

Background & Industry Context

- **Telecom industry faces intense competition and customer switching**
- **Customer churn can be influenced by pricing, service quality, and support experience**
- **Predictive analytics enables targeted retention offers and improved satisfaction**

Dataset Overview & Preparation

- **Dataset contains customer demographics, usage patterns, and plan details**
- **Target variable: Churn (Yes/No)**
- **Preprocessing steps: handled missing values, encoded categorical features, normalized numerical values**
- **Split into training (80%) and testing (20%) sets for evaluation**

EDA Highlights & Key Insights

- **Higher churn among customers with international plans**
- **Significant churn correlation with high customer service call frequency**
- **Higher day-time charges associated with churn**
- **Class imbalance noted between churn and non-churn groups**

Models Used

- **Logistic Regression**
- **Random Forest**
- **Gradient Boosting**
- **Support Vector Machine (SVM)**
- **K-Nearest Neighbors (KNN)**
- **XGBoost**
- **Stacked Ensemble (Random Forest + Logistic Regression + Gradient Boosting)**

Model Comparison & Performance

- **Metrics used: Recall, Precision, F1 Score, Accuracy, ROC-AUC**
- **Ensemble and tuned models outperformed single base models**
- **Best Model: Stacked Ensemble**
  - **Recall: 0.8889**
  - **F1 Score: 0.7960**
  - **ROC-AUC: 0.9215**
  - **Avg CV Recall: 0.8339**

Feature Importance & Interpretability

- **Key features influencing churn:**
    - **Number of customer service calls**
    - **International plan status**
    - **Total day charge**
- **Feature importance validated with SHAP/permutation tests for transparency**

Business Insights & Implications

- **Frequent support calls indicate dissatisfaction and higher churn risk**
- **International plan churners may need targeted pricing or service quality improvements**
- **High day-time charge customers may be retained through tailored offers**

Ethical Considerations

- **Ensure fairness and avoid bias toward specific customer segments**
- **Protect customer data privacy and comply with data regulations**
- **Transparent communication on how predictions are used for retention actions**

Future Work & Recommendations

- **Incorporate temporal trends and historical plan changes**
- **Explore deep learning for large-scale telecom datasets**
- **Deploy as a real-time churn prediction system with alerting capabilities**

Conclusion & Takeaways

- **Predictive modeling enables proactive churn management**
- **Stacked Ensemble approach provided the best performance**
- **Data-driven strategies can improve retention and reduce revenue loss**

THANK YOU