MOHAMED ALPHA

malpha@andrew.cmu.edu

MASTER OF SCIENCE IN ELECTRICAL AND COMPUTER ENGINEERING - [MSECE]

04-638: PROGRAMMING FOR DATA ANALYTICS

TELCO CUSTOMER CHURN ANALYSIS: A REFLECTION

14th December 2024.

Abstract

This project addressed the critical challenge of customer churn in telecommunications through advanced analytics and machine learning. The analysis combined unsupervised and supervised learning techniques to identify at-risk customers and develop targeted retention strategies, achieving high predictive accuracy (AUC > 0.96).

1. Solution Approach

1.1 Data Preprocessing & EDA

1. Identified bimodal tenure distribution (0-10 and 60-70 months)
2. Discovered strong correlation between tenure and total charges (0.82)
3. Found negative correlation between tenure and churn (-0.35)

1.2 Unsupervised Learning

1. Implemented K-means clustering
2. Used elbow method for optimal clusters
3. Applied t-SNE for dimensionality reduction

1.3 Supervised Learning

1. Split data: 70% training, 15% validation, 15% test
2. Implemented Logistic Regression, Random Forest, XGBoost
3. Addressed class imbalance using SMOTE
4. Evaluated using AUC-ROC, precision, recall, F1-score

1.4 Integration

1. Combined clustering with churn probabilities
2. Developed segment-specific strategies
3. Created visualization dashboard

2. Technical and Analytical Challenges

2.1 Data Processing Challenges

1. Bimodal distribution in tenure months (peaks at 0-10 and 60-70 months) required specialized handling
2. Complex feature correlations (Tenure-Total Charges: 0.82, Tenure-Churn: -0.35) demanded careful feature engineering

2.2 Modeling Challenges

1. Class imbalance in customer segments required SMOTE implementation
2. Feature selection complexity due to high correlation between certain variables

3. Ethical Considerations

The project identified significant algorithmic bias risks:

1. Algorithm bias
2. Demographic disparities in predictions (e.g., higher churn rates for seniors)
3. Privacy concerns

4. Mitigation strategies

1. Regular bias audits across demographic segments
2. Implementation of fairness-aware modeling techniques
3. Strict data privacy protocols

5. Future Improvements

1. Development of real-time prediction capabilities
2. Enhanced feature engineering incorporating temporal patterns
3. Dashboard development
4. Automated intervention trigger system

Conclusion

The integration of clustering with churn predictions enabled targeted retention strategies, offering a comprehensive solution to the telecommunications industry's churn challenges.

END