Audience - Questions and Answers

Predicting Customer Churn

Telecommunications Company

(Project 01 – Milestone 03)

Bellevue University
DSC680 – Applied Data Science

Submitted By:
Debabrata Mishra

Instructor:
Amirfarrokh Iranitalab

Here are ten common questions and answers that an audience might ask regarding the analysis of customer churn in the telecom industry:

1. **What factors were the most significant predictors of customer churn?**

   The key predictors of churn were month-to-month contracts, which had higher churn rates; shorter customer tenure; higher monthly charges; and lack of tech support or online security services, as customers without these services were more likely to churn.

2. **How did you handle missing values, and why did you choose that method?**

   Missing values in the 'Total Charges' column were addressed by removing rows with null values since they corresponded to customers with zero tenure, likely new customers who had not yet incurred charges. This ensured data integrity and accurate analysis.

3. **Why did you use SMOTE for balancing the dataset?**

   SMOTE (Synthetic Minority Over-sampling Technique) was employed to generate synthetic samples of the minority class (churned customers), helping to balance the dataset and improve the model's ability to predict churn by providing a more evenly distributed training set.

4. **What key insights did the analysis reveal about why customers churn?**

   The analysis showed that customers with month-to-month contracts, higher monthly charges, and no additional services such as tech support or online security were more likely to churn. This indicates that cost concerns, contract flexibility, and lack of added services are major drivers of churn.

5. **How did you ensure your models weren't overfitting the data?**

   Overfitting was mitigated through cross-validation, which assessed model performance on different data subsets to ensure generalization to unseen data. Hyperparameter tuning and performance metric monitoring were also used to optimize model parameters and reliability.

6. **Why were Random Forest and Gradient Boosting chosen as primary models?**

   Random Forest and Gradient Boosting were selected for their ability to handle complex data relationships. Random Forest provides high accuracy and interpretability, while Gradient Boosting offers strong predictive performance by combining multiple weak models into a robust predictor.

7. **What challenges did you encounter during data preprocessing, and how did you address them?**

   Challenges included managing missing values, ensuring data consistency, and converting categorical variables into numerical formats. These were resolved by cleaning the data, removing rows with missing values, creating dummy variables for categorical data, and standardizing numerical features.

8. **What are the limitations of your analysis, and how might they affect the results?**

   Limitations include potential bias from the dataset's sampling method, exclusion of external factors like competitor actions, and the static nature of historical data that may not fully capture evolving customer behavior. These limitations could impact the generalizability of the results.

9. **What recommendations can you provide to help reduce customer churn?**

   Recommendations include offering incentives for longer-term contracts to retain customers, bundling services with discounts for those paying higher monthly charges, enhancing customer support, promoting additional services like tech support, and implementing continuous monitoring to address churn risks proactively.

10. **What additional data would improve future churn prediction models?**

    To enhance future models, collecting data such as customer satisfaction scores, customer interactions with support services, usage patterns of telecom services, and competitor offerings would be beneficial. Tracking social media sentiment and customer feedback could also provide deeper insights into churn drivers.