Audience – Questions and Answers

Project1

Shanthibooshan Subramanian

Bellevue University

DSC680

Amirfarrokh Iranitablob

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

1. What specific factors did you find to be the most significant predictors of customer churn?

   Key predictors of customer churn were identified as the type of contract with month-to-month contracts having higher churn rates, tenure as customers with shorter tenure were more likely to churn, monthly charges as higher charges were linked to higher churn, and whether customers had tech support or online security services as those 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 'TotalCharges' column were handled by dropping rows with null values, as these rows also had zero tenure, indicating that they might be newly joined customers who haven't incurred charges yet. This method was chosen to maintain data integrity and ensure accurate analysis.

3. Can you explain the rationale behind using SMOTE for balancing the dataset?

   SMOTE (Synthetic Minority Over-sampling Technique) is used to address the class imbalance by generating synthetic samples of the minority class (churned customers). This approach was selected to improve the model's ability to predict churn by providing a more balanced training set, enhancing the detection of churned customers.

4. What are the key insights you discovered about why customers churn?

   The analysis revealed that customers with month-to-month contracts, higher monthly charges, and no additional services like tech support or online security were more prone

to churn. These insights suggest that contract flexibility, cost concerns, and lack of added services are critical factors driving customer churn.

5. How did you ensure that the models you built were not overfitting the data?

To prevent overfitting, cross-validation was employed to evaluate model performance across different subsets of the data, ensuring that the model generalizes well to unseen data. Hyperparameter tuning was also performed to optimize model parameters and avoid overfitting. Additionally, regular monitoring of performance metrics helped ensure model reliability.

6. Why did you choose Random Forest Classifier and Gradient Boosting as your primary models?

Random Forest and Gradient Boosting were chosen due to their robustness and effectiveness in handling complex data relationships. Random Forest offers high accuracy and interpretability, while Gradient Boosting provides strong predictive performance by combining multiple weak learners into a robust model. These models were selected for their ability to deliver high predictive accuracy and insights.

7. What were the main challenges you faced during data preprocessing and how did you overcome them?

Challenges included handling missing values, ensuring data consistency, and transforming categorical variables into numerical formats. These were overcome through careful data cleaning, dropping rows with null values, generating dummy variables for categorical data, and standardizing numerical features to prepare the data for modeling.

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

Limitations include potential biases due to the dataset's sampling method, the absence of external factors that might influence churn example competitor actions, market conditions, and the static nature of historical data which may not fully capture evolving

customer behaviors. These limitations could affect the generalizability of the results to broader contexts.

9. What actionable recommendations can you provide to the telecom company based on your analysis to help reduce customer churn?

Actionable recommendations to reduce customer churn include offering incentives for longer-term contracts to retain month-to-month customers, bundling services and providing discounts for those with higher monthly charges, improving customer support and promoting additional services like tech support and online security, and implementing continuous monitoring and feedback mechanisms to proactively address churn risks.

10. What additional data or features would you recommend collecting to improve the accuracy of future churn prediction models?

To improve the accuracy of future churn prediction models, I recommend collecting additional data such as customer satisfaction scores from surveys, details on customer interactions with support services, usage patterns of specific telecom services, and competitor service offerings. Additionally, tracking social media sentiment and customer feedback can provide deeper insights into customer sentiment and potential churn reasons