Lauren Butcher

July 21, 2024

DSC 680

**10 Questions an Audience Would Ask**

1. **How was the data collected and processed?**

* The data was collected from publicly available Kaggle datasets. The data includes customer demographics, account details, and service usage patterns. Data processing involved handling missing values, normalizing data, and feature engineering to create meaningful variables for analysis.

1. **What features were most important in predicting churn?**

* Using the Random Forest model, the most important features in predicting churn include:
  + Tenure
  + MonthlyCharges
  + Contract
  + TotalCharges
  + InternetService
  + PaymentMethod
  + TechSupport
  + OnlineSecurity
  + Dependents
  + PhoneService

1. **How did you handle missing data and outliers?**

* Missing data was handled by filling missing values with appropriate methods such as mean, median, or mode for numerical columns and the most frequent value for categorical columns. Outliers were detected and handled using techniques like removing or capping the outlier values.

1. **What metrics were used to evaluate model performance?**

* The metrics used to evaluate model performance include accuracy, precision, recall, F1-score, and AUC-ROC. These metrics provide a comprehensive evaluation of the model's predictive power and its ability to handle imbalanced datasets.

1. **How do you ensure the model is not biased?**

* To ensure the model is not biased, we implemented techniques such as stratified sampling during train-test split, feature importance analysis to ensure no discriminatory features are included, and regular monitoring of model predictions to check for biases.

1. **What are the potential impacts of incorrect predictions?**

* Incorrect predictions can lead to loss of customers who might have been retained with appropriate interventions. Additionally, false positives (predicting a customer will churn when they won't) can lead to unnecessary retention efforts and costs.

1. **How frequently should the model be retrained?**

* The model should be retrained periodically, such as monthly or quarterly, to incorporate new data and adapt to changing customer behavior patterns. Continuous monitoring and evaluation of model performance will help determine the optimal retraining frequency.

1. **How do you integrate the model with existing business systems?**

* The model can be integrated with existing business systems through APIs or batch processing systems. This allows the model to score customers in real-time or on a scheduled basis and provide actionable insights to customer retention teams.

1. **What are the costs associated with implementing the model?**

* The costs include data storage and processing, computational resources for model training and deployment, and costs associated with integrating the model into business systems. Additionally, there may be costs for ongoing monitoring and maintenance of the model.

1. **How do you measure the success of the retention strategies based on the model's predictions?**

* The success of retention strategies can be measured by tracking key metrics such as churn rate reduction, customer lifetime value, customer satisfaction scores, and the ROI of retention campaigns. Comparing these metrics before and after implementing the model-based strategies will help measure success.