CASE-BASED QUESTIONS

Case 1 : Customer Churn Prediction

Question :

You are provided with customer data for a telecom company, including demographic information, service usage, and whether the customer has churned or not. How would you build a model to predict customer churn?

Answer/Approach:

1. **Data Exploration:** Understand the data, check for missing values, and explore patterns.
2. **Feature Engineering:** Create relevant features like usage patterns, duration of service, and interaction with support.
3. **Model Selection:** Use models like logistic regression, decision trees, or ensemble methods like random forests or XGBoost.
4. **Evaluation:** Use metrics like accuracy, precision, recall, and AUC-ROC.
5. **Deployment:** Implement the model in a production environment and monitor performance.

CASE-BASED QUESTIONS

Case 2 : A/B Testing

Question :

An e-commerce company wants to test a new recommendation algorithm. How would you design an A/B test to measure its effectiveness?

Answer/Approach:

1. **Hypothesis Definition:** Clearly state the null and alternative hypotheses.
2. **Sample Size Calculation:** Determine the required sample size to achieve statistical significance.
3. **Randomization:** Randomly assign users to the control (current algorithm) and treatment (new algorithm) groups.
4. **Metrics:** Define success metrics such as click-through rate, conversion rate, and average order value.
5. **Analysis:** Use statistical tests to compare the performance of both groups.
6. **Conclusion:** Draw conclusions based on the results and make recommendations.

CASE-BASED QUESTIONS

Case 3 : Fraud Detection

Question :

You are tasked with detecting fraudulent transactions for a credit card company. How would you approach this problem?

Answer/Approach:

1. **Data Understanding:** Analyze transaction data to identify patterns indicative of fraud.
2. **Feature Engineering:** Create features such as transaction amount, frequency, location, and time of day.
3. **Modeling:** Use supervised learning models like logistic regression, decision trees, and anomaly detection methods like isolation forests.
4. **Evaluation:** Evaluate using metrics like precision, recall, F1 score, and confusion matrix.
5. **Monitoring:** Continuously monitor model performance and update the model as fraud patterns evolve.

CASE-BASED QUESTIONS

Case 4 : Sales Forecasting

Question :

A retail company wants to forecast sales for the next quarter. How would you approach this task?

Answer/Approach:

1. **Data Collection:** Gather historical sales data, including seasonal trends and external factors like holidays.
2. **Exploratory Data Analysis (EDA):** Identify patterns, trends, and anomalies in the data.
3. **Feature Engineering:** Create features such as moving averages, lagged values, and external indicators.
4. **Model Selection:** Use time series models like ARIMA, exponential smoothing, or machine learning models like random forests and gradient boosting.
5. **Evaluation:** Validate model performance using metrics like RMSE, MAE, and MAPE.
6. **Forecasting:** Generate forecasts and provide actionable insights.

CASE-BASED QUESTIONS

Case 5 : Recommender Systems

Question :

You need to build a recommendation system for an online streaming service. How would you approach it?

Answer/Approach:

1. **Data Understanding:** Analyze user behavior data, including watch history, ratings, and preferences.
2. **Collaborative Filtering:** Implement user-based or item-based collaborative filtering.
3. **Content-Based Filtering:** Use metadata like genre, actors, and directors to recommend similar content.
4. **Hybrid Approach:** Combine collaborative and content-based filtering for better recommendations.
5. **Evaluation:** Use metrics like precision, recall, and mean reciprocal rank (MRR) to evaluate the recommender system.
6. **Personalization:** Continuously update the model based on user interactions to improve recommendations.

CASE-BASED QUESTIONS

Case 6 : Sentiment Analysis

Question :

A company wants to analyze customer reviews to understand their sentiments about its new product. How would you proceed?

Answer/Approach:

1. **Data Collection:** Gather customer reviews from various sources like social media, websites, and surveys.
2. **Preprocessing:** Clean and preprocess the text data, including tokenization, stop-word removal, and stemming/lemmatization.
3. **Feature Extraction:** Use techniques like TF-IDF, word embeddings, or BERT for feature extraction.
4. **Modeling:** Use machine learning models like logistic regression, SVM, or deep learning models like LSTM and BERT.
5. **Evaluation:** Evaluate model performance using metrics like accuracy, precision, recall, and F1 score.
6. **Insights:** Analyze the results to provide actionable insights to the company.

Case 7 : Anomaly Detection

Question :

You are provided with server logs and need to detect anomalies in server performance. How would you approach this problem?

Answer/Approach:

1. **Data Understanding:** Analyze the server logs to identify normal and abnormal behavior patterns.
2. **Feature Engineering:** Create features like CPU usage, memory usage, request count, and error rates.
3. **Modeling:** Use unsupervised learning methods like clustering (e.g., DBSCAN), isolation forests, or autoencoders for anomaly detection.
4. **Evaluation:** Validate the model using techniques like ROC curve and precision-recall curves.
5. **Deployment:** Implement the model in a monitoring system to detect anomalies in real-time and alert the relevant teams.

CASE-BASED QUESTIONS

Case 8 : Image Classification

Question :

A healthcare company needs to classify X-ray images to detect pneumonia. How would you approach this problem?

Answer/Approach:

1. **Data Collection:** Gather a dataset of labeled X-ray images.
2. **Preprocessing:** Preprocess the images by resizing, normalization, and augmentation to increase the dataset size.
3. **Model Selection:** Use convolutional neural networks (CNN) architectures like ResNet, VGG, or transfer learning models.
4. **Training:** Train the model using cross-validation to avoid overfitting.
5. **Evaluation:** Use metrics like accuracy, precision, recall, F1 score, and AUC-ROC.
6. **Deployment:** Implement the model in a clinical setting, ensuring it integrates with existing systems and provides explainable results.

CASE-BASED QUESTIONS

Case 9 : Natural Language Processing (NLP)

Question :

A customer support system needs to automatically categorize incoming support tickets. How would you approach this problem?

Answer/Approach:

1. **Data Collection:** Gather a dataset of historical support tickets and their categories.
2. **Preprocessing:** Clean and preprocess the text data, including tokenization, stop-word removal, and stemming/lemmatization.
3. **Feature Extraction:** Use techniques like TF-IDF, word embeddings, or BERT for feature extraction.
4. **Modeling:** Use classification models like logistic regression, SVM, or deep learning models like LSTM and BERT.
5. **Evaluation:** Evaluate model performance using metrics like accuracy, precision, recall, and F1 score.
6. **Deployment:** Integrate the model into the support system to automatically categorize new tickets and continuously improve based on user feedback.

CASE-BASED QUESTIONS

Case 10 : Market Basket Analysis

Question :

A grocery store wants to analyze customer purchase patterns to increase sales. How would you approach this problem?

Answer/Approach:

1. **Data Collection:** Gather transaction data, including items purchased and transaction timestamps.
2. **Preprocessing:** Clean the data, removing any inconsistencies or missing values.
3. **Association Rule Mining:** Use algorithms like Apriori or FP-Growth to find frequent itemsets and generate association rules.
4. **Evaluation:** Evaluate the rules using metrics like support, confidence, and lift.
5. **Insights:** Analyze the results to identify patterns and provide recommendations to increase cross-selling and up-selling.
6. **Implementation:** Implement changes in the store layout, promotions, and marketing strategies based on the insights.