



**PLACEMENTDOST**

## Machine Learning Assessment questions

### Note:

Please be advised that these assessment inquiries are designed to accommodate interns with diverse skill levels, ranging from novices to seasoned analysts. Should you encounter any challenging questions, you are encouraged to seek solutions independently or reach out to us for assistance at [intern@placementdost.com](mailto:intern@placementdost.com). Best wishes for success in completing the assessment!

Please download dummy datasets for these questions from the internet!!

1. Detect Spam Emails

Question: Using Tensorflow tool, how can you detect spam emails automatically so the user's inbox remains clean? Perform EDA on dataset to detect spam mail.

2. Bitcoin Price Prediction

Question: Using numpy, matplotlib, sklearn libraries, how can you predict a signal that indicates whether buying a particular stock will be helpful or not by using ML.

3. Parkinsons disease detection

Question: Using numpy, matplotlib, sklearn libraries, how can you predict whether a person has Parkinson's disease or not?

## Additional Questions:

### Detect Spam Emails

#### 1. Feature Engineering for Email Text:

- In the context of spam email detection, what text-based features would you consider when preprocessing and engineering features from email content? How might these features contribute to model performance?

#### 2. Handling Imbalanced Datasets:

- Spam emails are often a minority class in datasets. How would you address class imbalance during training? Discuss methods such as oversampling, undersampling, or using different evaluation metrics in this scenario.

#### 3. NLP Techniques:

- Explain the natural language processing (NLP) techniques that can be applied to extract meaningful information from email text. How can techniques like tokenization and TF-IDF be used in the context of spam detection?

#### 4. Algorithm Selection:

- Compare and contrast the performance of different machine learning algorithms (e.g., Naive Bayes, Support Vector Machines, Decision Trees) for spam email detection. What factors influence your choice of algorithms?

#### 5. Evaluation Metrics:

- Discuss appropriate evaluation metrics for assessing the performance of a spam email detection model. How would you interpret metrics such as precision, recall, and F1-score in the context of spam detection?

### Parkinsons disease detection

#### 1. Feature Selection:

- Explain the process of feature selection for Parkinson's disease detection. What features are commonly considered, and how do they contribute to the model's accuracy?

#### 2. Data Preprocessing:

- Describe the essential steps in preprocessing the dataset for a Parkinson's disease detection model. How do you handle missing data, outliers, and ensure data quality?

#### 3. Algorithm Selection:

- Compare the performance of different machine learning algorithms for Parkinson's disease detection. What factors influence your choice of algorithms, and how do they handle the specific characteristics of medical datasets?

#### 4. Cross-Validation Techniques:

- Discuss the importance of cross-validation in evaluating the performance of your model. Explain the use of k-fold cross-validation and how it mitigates overfitting in the context of Parkinson's disease detection.

#### 5. Imbalanced Datasets:

- Parkinson's disease datasets often exhibit class imbalance. How would you address this issue during the training phase, and what metrics would you use to evaluate model performance?

## Bitcoin price prediction

#### 1. Feature Selection:

- What features would you consider relevant for predicting Bitcoin prices using machine learning? How might external factors such as market sentiment, trading volumes, or macroeconomic indicators impact feature selection?

#### 2. Time Series Data Preprocessing:

- Describe the challenges associated with preprocessing time series data for Bitcoin price prediction. How would you handle issues such as seasonality, trends, and non-stationarity in the data?

#### 3. Algorithm Selection for Time Series Prediction:

- Compare the suitability of different machine learning algorithms (e.g., ARIMA, LSTM, Random Forest) for time series prediction in the context of Bitcoin prices. What factors influence your choice of algorithm?

#### 4. Hyperparameter Tuning:

- Explain the role of hyperparameter tuning in optimizing the performance of a machine learning model for Bitcoin price prediction. What specific hyperparameters would you focus on, and how might they impact model performance?

#### 5. Handling Cryptocurrency Market Volatility:

- The cryptocurrency market is known for its volatility. How would you design a model that can adapt to sudden price fluctuations and maintain accuracy in predicting Bitcoin prices during both stable and volatile periods?