1. What are the Different dimensions of Big Data?
2. What are the differences between computational modeling and data modeling?
3. How to determine data quality?
4. Why do we need discretization of data?
5. What is the main challenge of discretization?
6. Different data cleaning steps?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_1. What do you mean by sample and population?

2. Why do you need confidence interval estimation instead of point estimation?

3. What is null hypothesis? How do we falsify the null hypothesis?

4. Describe any two EDA techniques.

5. What is the difference between EDA and classical data analysis method?

1. What do you mean by feature generation of a dataset?

2. What are the different methods of text feature generation/extraction?

3. What is the drawback of bag of word model? How do you tackle it?

4. How tf-idf is used to extract features from the documents?

5. Describe Template Matching method for image prediction.

1. What do you mean overfitting?

2. What is Regularization? Describe L2 regularization method.

3. Difference between deep learning and feature engineering approach.

4. For which type of dataset do we use decision tree classifier?

5. Why Random forest classifier performs better than decision tree classifier?

1. What is the Loss function of linear regression model?

2. How do you decide the degree of polynomial for fitting a good predictor model?

3. How do we choose an optimal model?

4. Describe the principle of gradient descent algorithm for minimizing the error?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_1.