Naive Approach:

1. What is the Naive Approach in machine learning?
2. Explain the assumptions of feature independence in the Naive Approach.
3. How does the Naive Approach handle missing values in the data?
4. What are the advantages and disadvantages of the Naive Approach?
5. Can the Naive Approach be used for regression problems? If yes, how?
6. How do you handle categorical features in the Naive Approach?
7. What is Laplace smoothing and why is it used in the Naive Approach?
8. How do you choose the appropriate probability threshold in the Naive Approach?
9. Give an example scenario where the Naive Approach can be applied.

KNN:

10. What is the K-Nearest Neighbors (KNN) algorithm?
11. How does the KNN algorithm work?
12. How do you choose the value of K in KNN?
13. What are the advantages and disadvantages of the KNN algorithm?
14. How does the choice of distance metric affect the performance of KNN?
15. Can KNN handle imbalanced datasets? If yes, how?
16. How do you handle categorical features in KNN?
17. What are some techniques for improving the efficiency of KNN?
18. Give an example scenario where KNN can be applied.

Clustering:

19. What is clustering in machine learning?
20. Explain the difference between hierarchical clustering and k-means clustering.
21. How do you determine the optimal number of clusters in k-means clustering?
22. What are some common distance metrics used in clustering?
23. How do you handle categorical features in clustering?
24. What are the advantages and disadvantages of hierarchical clustering?
25. Explain the concept of silhouette score and its interpretation in clustering.
26. Give an example scenario where clustering can be applied.

Anomaly Detection:

27. What is anomaly detection in machine learning?
28. Explain the difference between supervised and unsupervised anomaly detection.
29. What are some common techniques used for anomaly detection?
30. How does the One-Class SVM algorithm work for anomaly detection?
31. How do you choose the appropriate threshold for anomaly detection?

32. How do you handle imbalanced datasets in anomaly detection?
33. Give an example scenario where anomaly detection can be applied.

Dimension Reduction:

34. What is dimension reduction in machine learning?
35. Explain the difference between feature selection and feature extraction.
36. How does Principal Component Analysis (PCA) work for dimension reduction?
37. How do you choose the number of components in PCA?
38. What are some other dimension reduction techniques besides PCA?
39. Give an example scenario where dimension reduction can be applied.

Feature Selection:

40. What is feature selection in machine learning?
41. Explain the difference between filter, wrapper, and embedded methods of feature selection.
42. How does correlation-based feature selection work?
43. How do you handle multicollinearity in feature selection?
44. What are some common feature selection metrics?
45. Give an example scenario where feature selection can be applied.

Data Drift Detection:

46. What is data drift in machine learning?
47. Why is data drift detection important?
48. Explain the difference between concept drift and feature drift.
49. What are some techniques used for detecting data drift?
50. How can you handle data drift in a machine learning model?

Data Leakage:

51. What is data leakage in machine learning?
52. Why is data leakage a concern?
53. Explain the difference between target leakage and train-test contamination.
54. How can you identify and prevent data leakage in a machine learning pipeline?
55. What are some common sources of data leakage?
56. Give

an example scenario where data leakage can occur.

Cross Validation:

57. What is cross-validation in machine learning?
58. Why is cross-validation important?

59. Explain the difference between k-fold cross-validation and stratified k-fold cross-validation.
60. How do you interpret the cross-validation results?