Q:-1 Ans- D

Q:-2 Ans- A

Q:-3 Ans- C

Q:-4 Ans- C) 1 and 3

Q:-5 Ans- D) 1-3-2

Q:-6 Ans- B

Q:-7 Ans- C

Q:-8 Ans- B

Q:-9 Ans- B,C,D

Q:-10 Ans- A

Q-11 In which situation One-hot encoding must be avoided? Which encoding technique can be used in such a case?

Ans :- There are some situation where avoided one hot encoding .

High Cardinality

Correlated categories

Missing Values

Label Leakage

Q-12:-In case of data imbalance problem in classification, what techniques can be used to balance the dataset? Explain them briefly.
Ans:- There are several techniques that can be used to balance imbalanced datasets in classification problems, including undersampling, oversampling, combination of undersampling and oversampling, cost-sensitive learning, and ensemble learning. The choice of the appropriate technique depends on the specific dataset and the machine learning algorithm used.

Q:-13. What is the difference between SMOTE and ADASYN sampling techniques?

Ans:- SMOTE and ADASYN are both effective oversampling techniques for addressing class imbalance, but they differ in how they generate synthetic samples for the minority class. The choice of the appropriate technique depends on the specific dataset and the machine learning algorithm used.

Q:-14. What is the purpose of using GridSearchCV? Is it preferable to use in case of large datasets? Why or why not?

Ans:- GridSearchCV is a powerful technique for hyperparameter tuning that can be used for any size of the dataset. However, it can be computationally expensive for large datasets, and other techniques like RandomizedSearchCV or Bayesian optimization may be preferable in such cases.

Q:-15. List down some of the evaluation metric used to evaluate a regression model. Explain each of them in brief.

Ans:- There are several evaluation metrics used to evaluate the performance of regression models. Some of the most common ones are:

Mean Squared Error (MSE): MSE measures the average squared difference between the predicted and actual values. It is calculated by taking the average of the squared differences between the predicted and actual values of the target variable.

Root Mean Squared Error (RMSE): RMSE is the square root of the MSE. It measures the average distance between the predicted and actual values, in the same units as the target variable.

Mean Absolute Error (MAE): MAE measures the absolute difference between the predicted and actual values. It is calculated by taking the average of the absolute differences between the predicted and actual values of the target variable.

R-squared (R2): R2 measures the proportion of the variance in the target variable that is explained by the regression model. It ranges from 0 to 1, with higher values indicating better model performance.

Explained Variance Score (EVS): EVS measures the proportion of the variance in the target variable that is explained by the regression model. It ranges from 0 to 1, with higher values indicating better model performance.