1. B
2. A
3. B
4. D
5. D
6. B
7. C
8. C
9. B
10. C
11. When the number of categories in the dataset is quite large. One Hot Encoding should be avoided in this case as it can lead to high memory consumption.
12. Methods used to deal with unbalanced datasets –
    1. Random Undersampling and Oversampling - A widely adopted and perhaps the most straightforward method for dealing with highly imbalanced datasets is called resampling. It consists of removing samples from the majority class (under-sampling) and/or adding more examples from the minority class (over-sampling).
    2. Undersampling and Oversampling using imbalanced-learn. It provides a variety of methods to undersample and oversample –
        - Undersampling using Tomek Links : In this algorithm, we end up removing the majority element from the Tomek link, which provides a better decision boundary for a classifier.
        - In SMOTE (Synthetic Minority Oversampling Technique) we synthesize elements for the minority class, in the vicinity of already existing elements.
    3. Class weights in the models
    4. Change your Evaluation Metric
    5. Also :
        - Collect more data
        - Treat the problem as anomaly detection
13. The difference between ADASYN and SMOTE is that the ADASYN uses a density distribution as a criterion to automatically decide the number of synthetic samples that must be generated for each minority sample by adaptively changing the weights of the different minority samples to compensate for the skewed distributions. The SMOTE generates the same number of synthetic samples for each original minority sample.
14. GridSearchCV tries all the combinations of the values passed in the dictionary and evaluates the model for each combination using the Cross-Validation method. Hence after using this function we get accuracy/loss for every combination of hyperparameters and we can choose the one with the best performance.
    It is not preferrable to use for large datasets as its very time consuming.
15. There are 3 methods to evaluate regression model :
    1. Mean absolute error - MAE is a very simple metric which calculates the absolute difference between actual and predicted values.
    2. Mean squared error - It represents the squared distance between actual and predicted values. we perform squared to avoid the cancellation of negative terms and it is the benefit of MSE.
    3. Root mean squared error - It is a simple square root of mean squared error.