1. Ans) None of these
2. Ans) max_depth
3. A) SMOTE
4. B) 1 only
5. D) 1-3-2
6. D) Logistic Regression
7. Ans) ) None of the above
8. D) Lasso will cause some of the coefficients to become 0 A) Ridge will lead to some of the coefficients to be very close to 0
9. A) remove both features from the dataset C) Use ridge regularization
10. A) Overfitting C) Underfitting
11. we prefer one hot encoding when we require the values in binary form otherwise we must avoid the one hot encoding. Both the encoding technique perform excellent as per their own way ,mostly prefer to use Label encoding,when we have to deal with lots of data and we neet it in continuous values
12. In an imbalanced dataset first we need to think what we require to oversampling or downsampling as per that we need to use the sampling technique .There are different sampling techniques to use the most people prefer SMOTE to balance the data.
13. ) The key difference between ADASYN and SMOTE is that the former 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 data
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.
15. Mean Absolute Error, Mean Squared Error , Root Mean Square Error , Root Mean square Log Error ,these are the list of evaluation metrics that are mainly using this metrics to calculate the errors. Mean Absolute Error => MAE is sum of absolute error if we have 100 as actual value and 130 as a predicted value here the absolute error is 30 it doesn't consider the direction whether it is negative and positive values , we are calculating the all absolute error and finding the mean from it Mean Square Error => In mean square error we calculate the error occurred in between actual value and predicted value and squaring them in whether it is negative value it will automatically become the positive value and finding the mean from that all Root Mean Square Error => In this the formula is very similar to mean square error it is just we need to add the square to sign in it here it is indicates the residual error it is always positive and lower value indicates the better performance ideal value would be 0 but it is never achieved. Root Mean square Log Error => it is calculated at logarithmic scale ,RMSLE is added 1 as constant of actual and predicted value because they can be 0 but log of 0 will be undefined like we have actual value=100 and predicted value=130 we add 1 as constant in each 101 and 131 then we will find the log of actual and predicted value after that we will find the error from the value then squaring it after that we will calculate the mean