1. D) None of these

2. A) max\_depth

3. A) SMOTE

4. C) 1 and 3

5. A) 3-1-2

6. D) Logistic Regression

7. C)

8. A)

9. B) remove only one of the features

10. A) overfitting

12. Consider the case of a highly imbalanced dataset in which we have 90% data points belonging to label 1 and only 10% belonging to label 0. In that case even if we naively assign label 1 to all the data points we will get 90% accuracy because 90% data-points are of label 1 so, accuracy here will not be a good metric to evaluate the performance of a model. Even a very naive model like giving label 1 without even looking at any feature will give you 90% accuracy so, the accuracy will not be a good metric to use here.

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 distributions. The latter generates the same number of synthetic samples for each original minority sample.

14. **GridSearchCV** is a library **function** that is a member of sklearn's model\_selection package. It helps to loop through predefined hyperparameters and fit your estimator (model) on your training set. So, in the end, you can select the best parameters from the listed hyperparameters.

15. **Mean Squared Error:**MSE or Mean Squared Error is one of the most preferred metrics for regression tasks. It is simply the average of the squared difference between the target value and the value predicted by the regression model. As it squares the differences, it penalizes even a small error which leads to over-estimation of how bad the model is. It is preferred more than other metrics because it is differentiable and hence can be optimized better.

**Root Mean Squared Error:**RMSE is the most widely used metric for regression tasks and is the square root of the averaged squared difference between the target value and the value predicted by the model. It is preferred more in some cases because the errors are first squared before averaging which poses a high penalty on large errors. This implies that RMSE is useful when large errors are undesired.

**Mean Absolute Error:** MAE is the absolute difference between the target value and the value predicted by the model. The MAE is more robust to outliers and does not penalize the errors as extremely as mse. MAE is a linear score which means all the individual differences are weighted equally. It is not suitable for applications where you want to pay more attention to the outliers.

**R² Error:**Coefficient of Determination or R² is another metric used for evaluating the performance of a regression model. The metric helps us to compare our current model with a constant baseline and tells us how much our model is better. The constant baseline is chosen by taking the mean of the data and drawing a line at the mean. R² is a scale-free score that implies it doesn't matter whether the values are too large or too small, the R² will always be less than or equal to 1.

**Adjusted R²:**Adjusted R² depicts the same meaning as R² but is an improvement of it. R² suffers from the problem that the scores improve on increasing terms even though the model is not improving which may misguide the researcher. Adjusted R² is always lower than R² as it adjusts for the increasing predictors and only shows improvement if there is a real improvement.