**MACHINE LEARNING WORKSHEET-5**

1.ANS) D) NONE OF THESE

2.ANS) A) MAX\_DEPTH

4.ANS) C) 1 AND 3

5.ANS) D) 1-3-2

6.ANS) A) DECISION TREE

7.ANS) C) CART can only create binary trees (a maximum of two children for a node), and CHAID can create multiway trees (more than two children for a node)

8.ANS) C) Ridge will cause some of the coefficients to become 0 D) Lasso will cause some of the coefficients to become 0,

9.ANS) B) remove only one of the features

10.ANS) A) OVERFITTING

11.ANS) if the categorical data is ordinal, we can't use one hot encoding there we use label encoding

### 12.ANS) using the right evaluation metrics, resample the training set like under-sampling, over-sampling, use K-fold Cross-Validation in the right way, ensemble different resampled datasets, resample with different ratios, cluster the abundant class, then we will design our own model

13 ANS) 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

14.ANS) 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.it is prefers to do in large datasets because by doing grid search cv there we will get a suitable hyperparameter in the algorithm for the dataset for predictions

15.ANS) The various metrics used to evaluate the results of the prediction are

1) mean Squared Error 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.

2) root-mean-squared-error 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.

3) mean-absolute-error 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.

4) R² or Coefficient of Determination - 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

5) adjusted R² - Adjusted R² depicts the same meaning as R² but is an improvement of it, Adjusted R² is always lower than R² as it adjusts for the increasing predictors and only shows improvement if there is a real improvement.

**PYTHON WORKSHEET – 3**

1.ANS) D) Int (‘32’)

2.ANS) C) 4

3.ANS) B) (a\*\*b) % c

4.ANS) A) <class ‘type’>

5.ANS) C) 65

6.ANS) D) method

7.ANS) B) false

8.ANS) B) sometimes

9.ANS) A, C, D

10.ANS) A, C, D

### 