ASSIGNMENT-8

MACHINE LEARNING

Q1. C

Q2. A

Q3. A

Q4. B

Q5. D

Q6. B

Q7. C

Q8. A & D

Q9.C & D

Q10.A

Q11.

When the categorical features present in the dataset are ordinal i.e for the data being like Junior, Senior, Executive, Owner. When the number of categories in the dataset is quite large. One Hot Encoding should be avoided. Label Encoding can be use instead of One hot Encoding.

Q12.

Resampling (Oversampling and Undersampling)
When we are using an imbalanced dataset, we can oversample the minority class using replacement. This technique is called oversampling. Similarly, we can randomly delete rows from the majority class to match them with the minority class which is called undersampling.

Q13.

Adasyn is a improved version of Smote. What it does is same as SMOTE just with a minor improvement. After creating those sample it adds a random small values to the points thus making it more realistic.

Q14.

GridSearchCV is a technique for finding the optimal parameter values from a given set of parameters in a grid. It's essentially a cross-validation technique. Grid Search CV technique is not recommended for large-

size datasets because the complexity of Grid Search CV increases with an increase in the number of parameters in the param grid.

Q15.

There are three error metrics that are commonly used for evaluating and reporting the performance of a regression model; they are:

1)Mean Squared Error (MSE): Mean squared error states that finding the squared difference between actual and predicted value.

2) Root Mean Squared Error (RMSE): As RMSE is clear by the name itself, that it is a simple square root of mean squared error.

3)Mean Absolute Error (MAE) : MAE is a very simple metric which calculates the absolute difference between actual and predicted values.

STATISTICS WORKSHEET-8

Q1. B

Q2. B

Q3. D

Q4. B

Q5. C

Q6. D

Q7. B

Q8. A

Q9. B

Q10. C

Q11. A

Q12. D

Q13.

ANOVA in SPSS is used as the test of means for two or more populations.

Q14.

There are three primary assumptions in ANOVA: The responses for each factor level have a normal population distribution. These distributions have the same variance. The data are independent.

Q15.

The only difference is one-way ANOVA has one independent variable, while a two-way ANOVA has two.