Machine Learning Worksheet - 7

1) A) GridSearchCV and B) RandomizedCV
2) A) Random Forest
3) B) Regularization will decrease.
4) A) It regularizes the decision tree by limiting the maximum depth up to which a tree can be grown.
5) C) In case of classification problem, the prediction is made by taking mode of the class labels predicted by the component trees.
6) C) Both of them.
7) B) Bias will decrease, Variance will increase.
8) B) Model is overfitting.
9) **Gini Index** = 1-([40/100]^2+[60/100]^2)
   = 1- 52/100 = 0.48
   Entropy = - ([40/100 * log (40/100)] + [60/100 * log (60/100)])
   = - ([0.172] + [0.408])
   = - 0.58

10) Advantages of Random forest over Decision tree are:-
    a) Random forest algorithms avoid & prevents overfitting whereas in decision tree there is always a scope of overfitting.
    b) Random forest gives accurate & precise results whereas in decision tree the results are not accurate.

11) Feature scaling is normally used to transform the numeric features in a dataset to a standard range so that the performance of the machine learning algorithm improves.

    The two techniques used for scaling are:-

    a) Normalization
    b) Standardization

12)

13) No, if the dataset is imbalanced then the accuracy is not a good metric to measure the performance of the model because working with imbalanced data the minority class is our interest most of the time. So machine learning algorithms favor the larger class & sometimes ignore the smaller class if the data is highly imbalanced.

14) F-score is also called F1-score, is a measure of a model's accuracy on a dataset. It is a way of combining the precision & recall of the model & it is defined as harmonic mean of the model's precision & recall.

F-score formula:-

F1 = 2 * (precision*recall/precision + recall)

15) The fit () method helps in fitting the data into the model whereas, the transform () method will transform the dataset to proceed with further data analysis steps while, fit_transform () method will determine the parameters & transform the dataset.

Statistics WorkSheet – 7

1) The probability of getting 6 as outcome :-
   190/1402 = 0.135
   b)

2) The probability of getting a digit with unit place digit odd number i.e. 1, 3, 5, 7, 9 is:-

   (52+44+20+56+40)/400 = 0.53

   d)

3) The probability that the tyre will last more than 9000 miles is:-

   (375+445)/1100 = 0.745

   c)

4) The probability if we buy new tyre then it will last in the interval (4000-14000) miles is:-

   (260+375)/1100 = 0.577

   b)

5) The probability that the card is odd:-

   1 – 3/10 = 0.7

   d)

6) The probability that the card is even:-

   1 – 1/8 = 0.87

   d)

7) The probability that the number 6 has appeared at least on one of the die is:-

   The sum of the numbers appearing observed to be 7 is = 6

   The number of time the 6 appears at least in one of die is = 2

   Probability = 2/6 = 0.33

   c)

8) The A event that there is atleast one head & B event the die shows a number greater than 4 is:-

A = (H, T), (H, 1), (H, 2), (H, 3), (H, 4), (H, 5), (H, 6)

B = (H, 5), (H,6)

Now probability of event A is:-

1/4 + 6/12 = 3/4

Probability of event B is:-

2/12 = 1/6

The probability is:-

P(E) = (1/6)/(3/4) = 0.22

b)

9) The probability that the Ross being at one of the ends of the line is:-

P (E) = 4/6 = 0.66

a)

10) The probability that the both are girls given that at least one of them is girl:-

P (E) = 1/3 = 0.33

a)

11) a) 0.33

12) c) 0.78

13) a) 0.345

14) The probability of getting sum of two numbers is less than 4, provided that the two numbers found on two dice are different is:-

2/36 = 0.055

d)

15) The probability of getting head up is:-

1/6 + 1/6 +1/3 = 2/3

b)