MACHINE LEARNING:

1.RSSis the better measure of goodness of fit model in regression as lower the value ,it means low variations in the data.so that it can explain the data properly.

2.

RSS-it measures the variation in the error between observed data and model values.

TSS-It shows the amount of variationin the observed data.

R2-it is the absolute amt of variation, asa proportion of total variation.

R2=1-RSS/TSS

3.Regularisation refers to technique that are used to calibrate Machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting.

Hence it can reduce the error in test data set. It also discourges learning a complex /flexible model ,so avoid the risk of overfitting.

4. GINI index,calculates the amt of probability of a specific feature that is classified incorrectly ,when selected randomly..

It is measure in one of the methods in DT algorithm to decide the optimal spilt from a root node and subsequent spilts.

Lower the GINI IMPURITY,higher is the homogeneity of the node.

1. It helps to find out the root node,intermediate node, and leaf node to develop the decision tree.
2. IT is used by CART
3. IT reaches its min (0),when all cases in the node fall into a single node.

6,.ENSEMBLE methods:

It is the technique that aims at improving the accuracy of results in models by combining multiple models instead of using a single model.The combined models increase the accuracy of the results significantly.

It is used for regression and classification models, where they reduce bias and variance to boost the accuracy of models.

BAGGING:it increases the accuracy of models through DT,which reduces th variation to a large extent.The reduction of variance increases accuracy, elminiating overfitting.

7. BAGGING BOASTING

1.The simplest way of combining 1.A way of combining predicions that belong to the different types.

Prediction that belong to the same type.

2 aim to decrese variance, not bias. 2. Aim to deacrese the bias,not variance

3.each model receive equal weight. 3.Models are weighted acc to theirperformance

4 Each model iis build independently. 4.new models are influenced by previous built model

5 It tries to solve the over-fitting problem. 5.It tries to reduce bias.

8. OUT OF BAG ERROR is the average error for each calculated error using predictions from the trees that do not contain in their respective bootstrap sample.

This allow the random forest to be fit and validated whilst beng trained.

It is a method of measuring the prediction error of randon forest.

9. K FOLD CROSS VALIDATION.

IT is used to solve the problem of train-test-spilt.To evaluate the performance of a model.

It is simple to understand.It selects the K-value and based on this k-value ,we spilt the data.

The letter K refers to number of spilts or folds that are made in the dataset.

K FOLD CROSS VALIDATION Iis a method of evaluating the performance of a ML model by dividing the data set into k distinct folds or subsets and then training and evaluating the model k times ,each time using a different fold as the test set and removing folds as the training set.

10. Hyper-paramater tuning consists of finding a set of optimal hyperparameter values for alearning algorithm while applying the optimized algorthim to any data set.

SQL

1. SELECT\*

FROM movie

1. SELECT tiltle

FROM movie

WHERE(RUNTIME= longest)

3 SELECT title

FROM movie

WHERE(revenue=highest)

4 SELECT title

FROM movie

WHERE(budget=maximum)

5 SELECT title

FROM movie

SELECT gender\_id,person\_id,character\_name,cast\_order

FROM movie\_cast

WHERE( movie.movie\_id=movie\_cast.movie\_id)

6. SELECT MAX(country\_name)

FROM country

WHERE country.country\_id=production\_country.country\_id

AND production\_country.movie.id= movie.movie\_id

7 SELECT genre\_id as genre, genre\_name as genere\_name

FROM genre

WHERE genre.genre\_id=movie\_genre.genre.id

AND movie\_genre.movie\_id=movie.movie\_id

8

8 SELECT language.language\_name as LANGUAGE NAME, COUNT (movie.title) as TITLE

FROM language, movie

WHERE language.language\_id=movie\_language.language\_id

AND movie\_language.movie\_id= movie.movie\_id

9 SELECT movie.title,movie\_crew.COUNT(person.id) as CREW NO,movie\_cast.COUNT(person\_id)as cast no

FROM movie,movie\_crew,movie\_cast

WHERE movie.movie\_id=movie\_crew.movie\_id

AND movie.movie\_id=movie\_cast.movie\_id

10 SELECT TOP(10) title

FROM movie

GROUP BY popularity

ORDER BY popularity DESC

12 SELECT title

FROM movie

WHERE movie\_status= ‘rumored’

13 SELECT country.country\_name MAX(movie.revenue)

FROM country,movie

WHERE country\_id=’USA’

AND country.country\_id=production\_country.country\_id

14 SELECT TOP(20) title

FROM movie

ORDER BY budget DESC

STATISTICS

1.d

2. d

3 a

4.a

5.a

6.b

7a

8.d

9d