How do we split node in a decision tree, while splitting the node what do we look for : high entropy or low entropy

out of all the metrics of decision tree : what you have used in your project ? I have used entropy. Now ques will be why not gini index or chi sqaure test

merits demerits of decision tree, overfitting in decsion tree

how do you check overfitting in decision tree? Why do you use random forest when you have decision tree ? Merits of Random Forest

if random forest is used for avoiding Overfitting then how?

difference between bagging and random forest

how do we take subset of feature / columns in random forest?

how do we perform feature selection in Random forest?

suppose, we have 16 features out of that you have pass 4 to be your subset, so how your RF will gonna choose those as a subset

difference between r squared and adjusted r squared

why do we need to combine precision and recall in f1 score, what is the need of f1 score?

confusion matrix, precision, recall, roc auc curve, f1 score

what is xg boost classifier? Why do we call it extreme gradient boosting?What is extreme word mean in xg boost? What kind of regularization we used in xg boost ?

why did you switch to data science / data science journey, what you have learn till now

Second round focused on EDA/Analysis part.

Explain your capstone project. Explain your approach for EDA.

What were the insights from EDA?

What are DDL and DML?

What is the difference between UPDATE and MODIFY?

What are different joins in SQL?

Give an example for each SQL join.

Explain your Capstone project.

Why do we use Logistic Regression over Linear Regression in classification

What is a cost function? What is the cost function for Logistic Regression?

What is Chi-Square test? Explain the Null and Alternate hypotheses for it.

What is t-test, z-test? When do we use t-test over Z-test?

When do we use AnoVa over t-test?

What are the different evaluation metrics used for classification?

What is Precision? Recall? F1-score?

How does threshold management work? Give an example when Precision is given more priority than Recall?

What are the libraries used to import models like KNN, Decision Trees, Random Forest in Python?

What are different Ensemble techniques? How does ensemble work?

Explain the differences between Bagging and Boosting.

How to calculate IQR?

What is positive skewness and negative skewness?

What are the two types of regression?

What is multiple linear regression?

What is Logistic Regression?

What is p-value and give an example?

linear regression in detail

assumptions.. What test will you use to check and how do you overcome it

what is the performance matrix did you use

how much regression technique do you know

how would you decide what is tha best line of regression

difference between chi2 and anova
what is the difference between MLE and probability

how doed MLE works
how does it maximise and what does it maximize

how will you use MLE in regression

approact for your capstone projecy
what challenges did you face?

what is testing of estimation

what is null hypothesis for checking normality

what do you use to check normality withing residuals

what is vif
what you infer from vif

logical reasoning :
100 doors are open then you toggle the even number doors then you toggle every third door from first. Then you toggle every fourth from first door.

then how many doors are open and how many are closed???

How will you import multiple excel sheets in a data frame?

What are the different types of data types?

Difference between lists/ tuples/ dictionaries?

Working of logistic regression?

Evaluation metrics of regression/classification model?

To Know More

SUBSCRIBE TO MY YOUTUBE CHANNEL