ML Assignment

* Student Name: PARTH NITESHKUMAR PATEL
* Student Id : 190CS098
* Subject Code & Name: ICS3414 Machine Learning

Q Explain Random Forest Algorithm

=> Random Forest is a supervised learning algorithm

=> This algorithm is mayorly used in :-
(1) Classification Problems

(2) Regression Problems

=> Random Forest Algorithm can handle both : (1) Continuous Variables (Case of Regression)

(2) Categorical Variables (Case of classification)

=> Here, the forest it builds, is an ensemble of decision trees, usually trained with "bagging method"

=> Here, ensemble means combining multiple models.

⇒ Bagging means creating a different training set (subset) from sample training data with replacement and the final output is based on majority voting.

⇒ In simple term, Random forest constructs multiple decision trees and merges them together to get a more accurate and stable prediction.

→ Random forest add randomness to the model while growing the tree.

* Steps involved in the Algorithm :-

(1) In Random forest, n number of random records are taken from the data set, having k number of records.

(2) Individual decision trees are constructed for each sample.

(3) Each decision tree will generate an output.

(4) Final output is considered based on Majority Voting or Averaging.

* Important Hyperparameters used:-

* To increase predictive Power:

(1) n_estimators → no. of trees the algo. builds before averaging

(2) max_features → max. number of features Random forest considers splitting a node

(3) mini_sample_leaf → determines the minimum number of leaves required to split an internal node.

* To increase speed:-

(1) n_jobs → tells how many processors allowed to use.

(2) random_state → controls randomness of the sample

(3) oob_score → means out of bag; a cross validation method

⇒ This algorithm is used in a lot of different fields, like banking, stock market, medicine, e-commerce, etc.