Consulting & Analytics Club
IIT Guwahati

Summer Analytics 2022

A primer course on Data Science and Machine Learning

Week 4 Quiz (Graded)

Total points 28/30

Hope that you've gone through the course content for week-4 as well as the covered assignment before attempting the quiz.

- This form accepts the solution only once, so make sure you don't press the submit button accidentally. No requests will be entertained.
- Use the SAME email ID which you used for registering for Summer Analytics 2022.
- Please follow the honor code, which otherwise may lead to harsh actions being taken.

All the best :)

Email *

hridayagrawal0102@gmail.com

0 of 0 points

Name *

Hriday Agrawal

Enrolment ID *

This is a 5-digit number of the form XXXXX. It can be found in the enrolment confirmation mail sent to you.

41395

Are you from IIT Guwahati ? *

- ○ Yes
- ◉ No

If you are from IIT Guwahati , provide your roll no.

Graded Quiz 20 of 20 points

✓ Decision trees can handle 2/2

- ○ Medium dimensional data
- ◉ high dimensional data ✓

None of these

Low dimensional data

✓ Choose from the following that are not decision tree nodes 2/2

End Nodes

Chance Nodes

None of the Above ✓

Decision Nodes

✓ Which of the following are the advantage/s of Decision Trees? Tick all the correct options. 3/3

Possible Scenarios can be added

Use a white box model, If given result is provided by a model

Worst, best and expected values can be determined for different scenarios

All of the above ✓

✓ Which of the following algorithms is not an example of ensemble learning algorithm? 2/2

Random Forest

Adaboost

Extra Trees

Gradient Boosting

Decision Trees ✓

Feedback

Decision trees don't aggregate the results of multiple trees so it is not an ensemble algorithm.

✓ Random Forest has _____ as base learning models 1/1

multiple decision trees ✓

bagging

entropy

None of these

✓ Which of the following is/are true about Random Forest and Gradient Boosting ensemble methods? 3/3

- [x] Both methods can be used for classification task ✓
- [ ] Random Forest is use for classification whereas Gradient Boosting is use for regression task
- [ ] Random Forest is use for regression whereas Gradient Boosting is use for Classification task
- [x] Both methods can be used for regression task ✓

✓ Machine Learning technique that helps in detecting the outliers in data. 2/2

- ( ) Clustering
- ( ) Classification
- (•) Anomaly Detection ✓
- ( ) All of the above

✓ Which of the following is true about the Gradient Boosting trees? 2/2

- ( ) In each stage, introduce a new regression tree to compensate the shortcomings of existing model
- ( ) We can use gradient decent method for minimize the loss function
- ( ) Neither A nor B
- (•) Both A and B ✓

✓ True or false: Traditionally, XGBoost is slower than lightGBM but it achieves faster training through the Histogram binning process. 1/1

- (•) True ✓
- ( ) False

✓ Which of the following is true about Naive Bayes ? 2/2

- ( ) assumes that all the features in a dataset are equally important
- ( ) assumes that all the features in a dataset are independent

None of the above

Both A and B ✓

Questions Based on Graded Assignment — 8 of 10 points

✗ Most important feature for XGBClassifier — 0/1

safety

maint

persons ✗

buying

Correct answer

safety

✓ Least Important Feature for XGBClassifier - — 1/1

maint

buying

doors ✓

lug_boot

✗ Now find out the class which had the highest Recall- for XGBClassifierwith parameters n_estimators=1000, learning_rate=0.1 Hint for both these questions: sklearn.metrics.classification_report . — 0/1

acc ✗

vgood

unacc

good

Correct answer

unacc

✓ You found the classification matrix. Now find out the class which had the lowest Precision-for XGBClassifierwith parameters n_estimators=1000, learning_rate=0.1 — 1/1

vgood

unacc

acc

good ✓

✓ Score for XGBClassifier with hyperparameters: n_estimators=1000 & learning_rate = 0.01. 1/1

- 0.97
- 0.94
- 0.98
- 0.95 ✓

✓ Score for Random Forest with hyperparameters: n_estimators=1000, max_depth=10, random_state=0 1/1

- 0.95 ✓
- 0.98
- 0.88
- 0.94

✓ Score for AdaBoost with hyperparameters: n_estimators=1000 & random_state=0 1/1

- 0.95
- 0.98
- 0.67
- 0.77 ✓

✓ Score for Decision Tree Classifier with criterion entropy index,max_depth=8, random_state=0. 1/1

- 0.95
- 0.92 ✓
- 0.77
- 0.98

✓ Gini impurity of root node in the decision - trees. Hint - see Visualize decision-trees. 1/1

- 0.557
- 0.657
- 0.357
- 0.457 ✓

✓ What is the leaf weight of XGBClassifier for criteria - person <1 (Yes). 1/1

- 0.0066
- -0.0066 ✓
- -0.066
- 0.066

You've reached the end of the quiz 0 of 0 points

I have read all my answers and this is my final submission. *

- YES