

Course Assessment Test

Course Title	Machine Learning with Python	Date	June 2024
Name		Dept	

- 1) This assessment test is to be given out before course commencement. Answers are to be filled in column entitled "Pre-Course Answer"
- 2) At the end of the course, the same assessment sheet is to be given out where answers are to be filled in column entitled "Post-Course Answer". Instructor will then share the answers and participants need to total the score in both "Pre" and "Post" columns through self-marking.
- 3) Assessment sheets will be collected for filling.

No	Question	Pre-Course Answer	Post-Course Answer
1	<p>Identify which of the following activities that machine learning algorithms / models are typically used in?</p> <ul style="list-style-type: none"> i. Descriptive analytics involving aggregating datasets for dashboard visualization ii. Predictive analytics for potential problems on the assembly line in manufacturing processes iii. Product recommendation for customers on an e-commerce website iv. Fraud detection for online transactions and trading by banks and stock exchanges <ul style="list-style-type: none"> a) Items i, ii and iii b) Items i, ii and iv c) Items ii, iii and iv d) All of the above items 		C
2	<p>In supervised learning, a labelled dataset is provided as input to a machine learning algorithm in order to train or fit the machine learning (ML) model. Identify the statement that best describes the ML model.</p> <ul style="list-style-type: none"> a) The ML model is a mathematical function that best approximates the relationship between independent variables (IV) and dependent variable (DV) of the labelled dataset b) The ML model identifies the statistical relationship between important contributory variables amongst the independent variables (IV) c) The ML model provides an estimate of the comparative variance in the dependent variable (DV) of the labelled dataset d) The ML model generates predictions for new data points through a statistical analysis of the dependent variable over a period of time (time series analysis) 		A
3	<p>In the fitting / training of the ML model, what role does the error / loss function play ?</p> <ul style="list-style-type: none"> a) The error function guides the ML algo to decide which independent variables contribute the most to the variance in the dependent variable b) The error function guides the ML algo to incrementally improve model accuracy by 		B

Course Assessment Test

	<p>changing the model parameters to minimize the error / loss function</p> <p>c) The error function guides the ML algo to identify the range of new data point values that are most likely to cause errors when predictions are generated using them</p> <p>d) The error function guides the ML algo to eliminate statistical outliers from anomalies in the dataset which in turn improves the overall predictive accuracy of the final ML model</p>		
4	<p>Which type of machine learning approach is used to predict a categorical variable value ?</p> <p>a) Regression</p> <p>b) Clustering</p> <p>c) Reinforcement learning</p> <p>d) Classification</p>		D
5	<p>Which of the following algorithms / models are generated in, or used for, supervised learning?</p> <p>i. Linear / Polynomial Regression</p> <p>ii. Decision Trees</p> <p>iii. Random Forest</p> <p>iv. Naïve Bayes</p> <p>a) Items i, ii and iii</p> <p>b) Items i, ii and iv</p> <p>c) Items ii, iii and iv</p> <p>d) All of the above items</p>		D
6	<p>A typical machine learning workflow involves a variety of key phases and activities. Which are the 2 phases / activities that likely to consume the most time in a typical workflow?</p> <p>i. Identifying the correct ML algo / model</p> <p>ii. Collecting datasets</p> <p>iii. Cleaning and exploring datasets</p> <p>iv. Refining / finetuning a selected ML algo / model</p> <p>a) Items i) and ii)</p> <p>b) Items i) and iii)</p> <p>c) Items ii) and iii)</p> <p>d) Items ii) and iv)</p>		C
7	<p>Identify the primary reason to perform a train - test split on the original dataset prior to training / fitting the ML model, instead of using the original dataset in its entirety</p> <p>a) Train - test split prevents the problem of overfitting whereby the generated model appears to be very accurate when tested with the same dataset used to train it, but performs poorly in generating accurate predictions for new samples</p> <p>b) Train - test split helps to improve the accuracy of the fitted model when the dataset size is very small</p> <p>c) Train - test split ensures that all significant classes of datapoints in the original datapoint are present in both training and testing to prevent bias in the generated</p>		A

Course Assessment Test

	<p>model</p> <p>d) Train - test split forces the data scientist to perform proper exploratory statistical data analysis on the dataset to identify which data points are best suited for fitting the model and which ones are best suited for testing instead</p>		
8	<p>Identify the common evaluation metrics for ML classification models from the list below</p> <ul style="list-style-type: none"> i. Accuracy ii. Confusion matrix iii. Category correctness iv. Recall <p>a) Items i, ii and iii b) Items i, ii and iv c) Items i, iii and iv d) All of the above items</p>		B
9	<p>Identify the common evaluation metrics for ML regression models from the list below</p> <ul style="list-style-type: none"> i. R-Squared ii. Mean Squared Error (MSE). iii. Root Mean Squared Error (RMSE). iv. Mean Absolute Error (MAE) <p>a) Items i, ii and iii b) Items i, ii and iv c) Items ii, iii and iv d) All of the above items</p>		D
10	<p>Identify the statements that correctly describe hyperparameter tuning as an approach to improving model accuracy</p> <ul style="list-style-type: none"> i. They are configuration variables that are internal to the model and their values are incrementally estimated from the training dataset by the ML algo during the training / fitting process ii. Each particular ML algo/model has a unique set of hyper parameters associated with it. iii. A specific combination of hyperparameters produces a model that is typically evaluated using the validation dataset iv. They can be set manually by the data scientist prior to the start of the training process <p>a) Items i, ii and iii b) Items i, ii and iv c) Items ii, iii and iv d) All of the above items</p>		C

Course Assessment Test



Total		
-------	--	--