Applied Data Science - Quiz 1

Name *

Kamalika P

Registered Roll Number/Registered Number *

Please enter the number as displayed in the profile section in the platform

913119104041

Registered Email id *

Please enter the email id used to login to the platform

kmlika22001@gmail.com

Select your Training Batch *

B7-1A3E ▼

Data Science is a ?

○ Branch of statistics

○ Branch of computer science

○ Branch of artificial intelligence

◉ Interdisciplinary, made up of all of the above

Machine learning is a subset of ?

○ Data Science

○ Deep Learning

◉ Artificial Intelligence

○ None of the above

Who developed Python Programming Language ?

○ Wick van Rossum

○ Wick van Rossum

◉ Guido van Rossum

○ Rasmus Lerdorf

○ Niene Stom

Which of the following is the correct extension of the Python file ?

◉ .py

○ .pl

○ .python

○ .p

What will be the output of the following Python code snippet?

```
for i in [1, 2, 3, 4][::-1]:
print (i)
```

◉ 4 3 2 1

○ error

If their is missing values in Categorical column, how do deal with?

○ Replace with Mean value

○ Replace with Median

○ Replace with Mode value

◉ Replace Quantile value

What is the use of the zeros() function in Numpy array in python ?

○ To make a Matrix with first row 0

○ To make a Matrix with all element 0

○ To make a Matrix with all diagonal element

◉ None of the above

What is a correct syntax to return the first row in a Pandas DataFrame?

○ df.get(0)

◉ df.loc[0]

○ df[0]

○ df

How do you get the list of columns that have null values in a DataFrames?

⦿ df.isnull().sum()

○ df.isnull().any()

○ df.isnull()

○ df.info()

To check the distribution of Numeric column ,which type of graph we will use ?

○ Box plot

⦿ Scatter plot

○ Dist plot

○ heat map

Applied Data Science - Quiz 2

Name *

Kamalika P

Registered Roll Number/Registered Number *

Please enter the number as displayed in the profile section in the platform

913119104041

Registered Email id *

Please enter the email id used to login to the platform

kmlika22001@gmail.com

Select your Training Batch *

B7-1A3E

Which of the following is a branch of statistics?

○ Descriptive statistics

○ Inferential statistics

○ Industry statistics

◉ Both A and B

The runs scored by a batsman in 5 ODIs are 31,97,112, 63, and 12. The mean is

○ 112

◉ 63

○ 97

○ 12

Which of the following is not a visualization under matplotlib

○ Scatter plot

○ Histogram

○ Box plot

◉ Table plot

To handle the null values in your data which method is used?

○ isnull()

◉ isfull()"

○ fillna()

○ fullna()

Which of the following sklearn method is used to split the data into training and testing

○ test_split()

◉ train_test_split()

○ test_train_split()

○ split_train_test()

Which of the following is correct to import the matplotlib library

○ import plt.matplotlib as plot

○ import plot.matplotlib as pt

○ import matplotlib.pt as plot

○ import matplotlib.pyplot as plt"

Which method is used for encoding the categorical variables

○ One Hot Encoder

○ Category Encoder

○ Label Encoder

◉ Both A & C

Which statistical method we can use for replacing missing values for categorical feature

○ Mean

○ Median

◉ Mode

○ All of the above

Which function we use to get descriptive statistics for the data

◉ description()

◯ describe()

◯ statistics()

◯ None of the above

Identify the outlier for the given data
34,23,43,28,56,125,52,40,26,24,37,25

◯ 56

◉ 23

◯ 125

◯ 34

Applied Data Science - Quiz 3

Name *

Kamalika P

Registered Roll Number/Registered Number *

Please enter the number as displayed in the profile section in the platform

9113119104041

Registered Email id *

Please enter the email id used to login to the platform

kmlika22001@gmail.com

Select your Training Batch *

B7-1A3E

Linear Regression is a machine learning algorithm based on *

○ unsupervised learning

◉ supervised learning

○ reinforcement learning

○ none of these

Regression models a target prediction value based on *

○ dependent variable

◉ independent variables

○ independent value

○ dependent value

Regression technique finds out a linear relationship between x (input) and
y(output) hence it is called as *

○ Hypothesis function

○ Related regression

● Linear Regression

○ none of these

Which Machine Learning technique use for dealing Categorical data? *

● Regression

○ Classification

○ Clustering

○ All of the above

How do you choose the root node while constructing a Decision Tree? *

◯ "An attribute having high entropy

◯ "An attribute having largest information gain

◯ "An attribute having high entropy and Information gain

⦿ None of the Mentioned

Choose a disadvantage of decision trees among the following. *

◯ Decision trees are robust to outliers

◯ Factor analysis

⦿ Decision trees are prone to overfit

◯ none of these

What is the term known as on which the machine learning algorithms build a model based on sample data? *

- ○ Data training
- ● Training data
- ○ Transfer data
- ○ None of the above

Machine learning is a subset of which of the following. *

- ● Artificial Intelligence
- ○ Deep learning
- ○ NLP
- ○ None of the above

The father of machine learning is *

(●) Geoffrey Everest Hinton

(○) Geoffrey hill

(○) Geoffrey chaucer

(○) Micheal Geoffrey

Suppose you got a training accuracy of 90% and a test accuracy of 50%. What *
happened with your model

(○) The model was over fitted with the training data

(○) The model was under fitted with the training data

(●) The model is absolutely fine

(○) None of the above

Applied Data Science - Quiz 4

Name *

Kamalika P

Registered Roll Number/Registered Number *
Please enter the number as displayed in the profile section in the platform

913119104041

Registered Email id *
Please enter the email id used to login to the platform

kmlika22001@gmail.com

Select your Training Batch *

B7-1A3E ▼

KNN Algorithm can be used for *

○ Only for Classification

○ Only for Regression

◉ Both Classification and Regression

○ None of the Mentioned

The goal of clustering a set of data is to *

○ divide them into groups of data that are near each other

○ choose the best data from the set

○ determine the nearest neighbors of each of the data

◉ predict the class of data

The most commonly used measure of similarity is the _____ or its square. *

◉ euclidean distance

○ city-block distance

○ Chebychev's distance

○ Manhattan distance

Which of the following is required by K-means clustering? *

○ defined distance metric

○ number of clusters

○ initial guess as to cluster centroids

◉ All of the above

which of the following is an ibm cloud service *

○ IBM watson

○ IBM NODRED

○ AUTO-AI

◉ All of the above

Which cross validation method use for hyperparameter tuning in machine learning *

○ Grid-searchcv

○ Randomisedsearchcv

○ BOTH

◉ None of the above

Techniques used for handling imbalanced dataset *

○ SMOTE

◉ Option OVERSAMPLING

○ UNDERSAMPLING

○ All of the above

Flask is a web development framework created in _____ language *

○ C

◉ Python

○ Java

○ Javascript

What is WSGI? *

- ⦿ Web Server Gateway Interface
- ◯ Web Service Gateway Interface
- ◯ Website Server Gateway Interface
- ◯ Web Static Gateway Interface

What is the purpose of the Kernel Trick? *

- ◯ To transform the problem from supervised to unsupervised learning.
- ◯ To transform the problem from regression to classification
- ⦿ To transform the data from nonlinearly separable to linearly separable
- ◯ To transform the problem from supervised to unsupervised learning