DataScience Specialization Test Series - Test 2 - Batch A & B - 11th October 2023

Total points 38/60

The respondent's email (**rajendraku0000@gmail.com**) was recorded on submission of this form.

0 of 0 points

Name *

Rajendra Kumar

DataScience Specialization Test Series - Test 2 - Batch A & B - 11th October 2023

38 of 60 points

✓ What is the output of the given code 1/1

```
for i in range(101,115):
if i == 103:
break
print(i)
```

- (●) 101 102 ✓
- ( ) 102
- ( ) 101 102 103
- ( ) None of the above

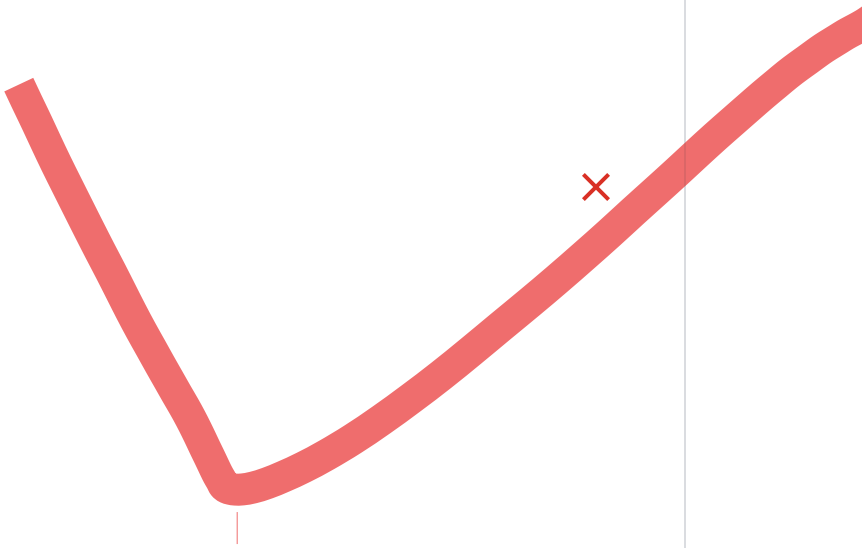
✗ What is the value of x

```
d = {'a': 1, 'b': 2, 'c': 3}
x = d.pop('b')
```

- (●) 2
- ( ) [1,3]
- ( ) {'a': 1, 'c': 3}
- ( ) None of the above

Correct answer

- (●) [1,3]



✓ **What is the purpose of dimensionality reduction techniques like Principal Component Analysis (PCA)?** 1/1

- ○ To increase the number of features in the dataset
- ○ To decrease the computational complexity of the model
- ◉ To eliminate irrelevant features and reduce data dimensionality ✓
- ○ To add noise to the data

✗ When performing matrix multiplication using np.dot or @ operator, what is the requirement for the inner dimensions of the matrices? 0/1

- ○ The inner dimensions must be equal.
- ○ The inner dimensions must be different.
- ◉ There is no specific requirement for the inner dimensions. ✗
- ○ "The inner dimensions must be integers.

Correct answer

- ◉ The inner dimensions must be equal.

✓ **Which of the following is an example of an unsupervised learning algorithm?** 1/1

- ○ Linear Regression
- ◉ K-Means Clustering ✓
- ○ Decision Trees
- ○ Support Vector Machines

✓ **In machine learning, what is the term "overfitting" referring to?** 1/1

- ◉ Model memorizes training data but fails to generalize ✓
- ○ Model's inability to learn from training data
- ○ Model's inability to memorize training data
- ○ Model's ability to generalize well

✓ What is the output of the given code 1/1

```
s = "My name is Nick"
x = s.count('n')
print(x)
```

- ◉ 1 ✓
- ○ 2
- ○ 0
- ○ TypeError

✓ What Keyword we use for creating an alias name? 1/1

- (●) as ✓
- ( ) with
- ( ) like
- ( ) in

✗ **Which of the following is an advantage of using an ensemble learning method like Random Forest over a single decision tree?** 0/1

- ( ) Random Forest is less computationally intensive.
- ( ) Random Forest is immune to overfitting.
- (●) Random Forest provides more interpretable results. ✗
- ( ) Random Forest reduces variance and improves generalization.

Correct answer

- (●) Random Forest reduces variance and improves generalization.

✗ What is the output of the given code 0/1

```
s = {1, 2, 3}
t = {2, 3, 4}
x = s ^ t

print(x)
```

- ○ {2,3}
- ○ {1,4}
- ◉ {1,2,3,4} ✗
- ○ {}

Correct answer

- ◉ {1,4}

✓ What method to use to all characters in the string are numbers? 1/1

- ◉ isdigit() ✓
- ○ isnum()
- ○ isnumeric()
- ○ notalpha()

✓ **Which evaluation metric is commonly used for binary classification problems when the dataset is imbalanced?** 1/1

- ○ Mean Absolute Error (MAE)
- ○ Mean Squared Error (MSE)
- ◉ F1 Score ✓
- ○ R-squared (R2)

✓ **What is the purpose of regularization techniques such as L1 and L2 regularization in machine learning?** 1/1

- ○ To increase model complexity
- ○ To reduce model bias
- ◉ To prevent overfitting by adding a penalty to the model's parameters ✓
- ○ To speed up model training

✓ What is the result of the following NumPy code? 1/1

```
import numpy as np

arr = np.array([1, 2, 3, 4, 5])
result = np.percentile(arr, 25)
```

- (●) 1 ✓
- ( ) 2.5
- ( ) 3
- ( ) 4

✓ **What is the loss function used in logistic regression for binary classification?** 1/1

- ( ) Mean Absolute Error (MAE)
- ( ) Mean Squared Error (MSE)
- (●) Cross-Entropy Loss (Log Loss) ✓
- ( ) Hinge Loss

✗ You have a NumPy array arr with shape (3, 4, 5). What is the result of the following operation? 0/1

result = np.mean(arr, axis=(0, 2))

- (●) A 4x3 array ✗
- ( ) A 3x5 array
- ( ) A 1D array with 4 elements
- ( ) A 1D array with 5 elements

Correct answer

- (●) A 1D array with 4 elements

✓ What is the result of the given code 1/1

import pandas as pd

df.xyz.map(dict(yes=1, no=0))

- (●) For the column xyz replaces yes to 1 and no to 0 ✓
- ( ) For the column xyz creates a dictionary where yes maps to 1 and no to 2
- ( ) SyntaxError
- ( ) TypeError

✓ **Which algorithm is commonly used for natural language processing (NLP) tasks, such as text classification and sentiment analysis?** 1/1

- ○ Random Forest
- ○ K-Means Clustering
- ◉ Naive Bayes ✓
- ○ Principal Component Analysis (PCA)

✓ **What is the main drawback of the k-nearest neighbors (KNN) algorithm?** 1/1

- ◉ It is sensitive to outliers. ✓
- ○ It cannot handle categorical data.
- ○ It requires a lot of training data.
- ○ It is computationally efficient for large datasets.

✗ Given a 2D NumPy array arr, how can you calculate the determinant of the submatrix formed by removing the first row and last column? 0/1

- ○ determinant = np.linalg.det(arr[1:, :-1])
- ◉ determinant = np.det_submatrix(arr, rows=[0], cols=[-1]) ✗
- ○ determinant = np.remove_row_last_col_det(arr)
- ○ determinant = np.linalg.det(arr[1:, 1:])

Correct answer

- ◉ determinant = np.linalg.det(arr[1:, :-1])

✓ What is the output of the given code 1/1

```
import pandas as pd

data = {'Name': ['nicki', 'hary', 'naman', 'jivit'],
'Age': [28, 24, 35, 32]
}
df = pd.DataFrame(data)
print(df['Age'].avg())
```

- (●) 29.75 ✓
- ( ) 28
- ( ) 30
- ( ) None of the above

✓ **What is the purpose of the term "dropout" in deep neural networks?**

- ( ) To remove outliers from the dataset
- (●) To prevent overfitting by randomly deactivating neurons during training ✓
- ( ) To reduce the learning rate of the network
- ( ) To increase the model's complexity

✗ What is the output of the given code 0/1
print('1234444'>'2111')

- (●) 1234444'>'2111' ✗
- ( ) TRUE
- ( ) FALSE
- ( ) 2111

Correct answer

- (●) TRUE

✓ What is the output type of the given code 1/1
string="It might rain today"

print(string.split('t'))

- ( ) string
- (●) list ✓
- ( ) set
- ( ) tuple

✗ What is the output of the given code 0/1

```
import pandas as pd
import numpy as np

df['col'] = df['col'].replace('-', np.nan)
df = df.dropna(axis=0, subset=['col'])
```

- ○ df replaces all rows with col value as'-' with 0
- ○ df drops all rows with col values as'-'
- ◉ df replaces all rows with col values as'-' with np.nan ✗
- ○ None of the above

Correct answer

- ◉ df drops all rows with col values as'-'

✓ **Which of the following is a hyperparameter of the k-nearest neighbors (KNN) algorithm?** 1/1

- ◉ Number of neighbors (k) ✓
- ○ Learning rate
- ○ Depth of the decision tree
- ○ Number of features

✓ **Which type of machine learning task involves predicting a continuous value, such as house prices or stock prices?** 1/1

- ○ Classification
- ○ Clustering
- ◉ Regression ✓
- ○ Reinforcement Learning

✗ **What is the purpose of the term "early stopping" in the context of training neural networks?** 0/1

- ○ To stop training as soon as a specific layer in the network is reached
- ○ To terminate training when the loss on a validation set starts to increase
- ◉ To stop training after a fixed number of epochs, regardless of performance ✗
- ○ To halt training if the learning rate becomes too large

Correct answer

- ◉ To terminate training when the loss on a validation set starts to increase

✓ **What is the primary difference between k-means clustering and hierarchical clustering?** 1/1

- ○ K-means is a supervised learning technique, while hierarchical clustering is unsupervised.
- ◉ K-means requires the number of clusters (k) to be specified in advance, while hierarchical clustering does not. ✓
- ○ . K-means always produces a dendrogram, while hierarchical clustering does not.
- ○ K-means is only applicable to binary classification problems.

✓ **In ensemble learning, what does bagging (Bootstrap Aggregating) involve?** 1/1

- ○ Training multiple models on the same dataset
- ○ Combining the predictions of multiple models
- ○ Reducing the complexity of a single model
- ◉ Randomly selecting subsets of the dataset for training ✓

✓ **Which optimization algorithm is commonly used to train deep neural networks?** 1/1

- ◉ Gradient Descent ✓
- ○ K-Means Clustering
- ○ Principal Component Analysis (PCA)
- ○ Random Forest

✗ You are given a 2D NumPy array arr of shape (4, 4). How can you obtain a new 2D array submatrix containing only the elements from rows 1 to 3 (inclusive) and columns 2 to 3 (inclusive)? 0/1

- (●) submatrix = arr[1:3, 2:3] ✗
- ( ) submatrix = arr[1:4, 2:4]
- ( ) submatrix = arr[0:2, 1:3]
- ( ) submatrix = arr[2:3, 1:3]

Correct answer

- (●) submatrix = arr[1:4, 2:4]

✗ What is the value of x
d = {'a': 1, 'b': 2, 'c': 3}
x =d.append('d':4) 0/1

- ( ) ('d':4)
- ( ) {('d':4})
- ( ) TypeError
- (●) AttributeError ✗

Correct answer

- (●) TypeError

✓ What is the output of the given code 1/1

```
string="It might rain today"
x=string[::-1]
print(x)
```

- (●) yadot niar thgim tI ✓
- ( ) it might rain today
- ( ) today rain might it
- ( ) TypeError

✓ What is the output of the given code 1/1

```
y = {2:20,4:40,6:60,8:80}
x = sum(y.values())

print(x%2)
```

- ( ) 100
- ( ) 50
- (●) 0 ✓
- ( ) 2

✓ What is the output of the given code

```
d = {'a': {1: 'one', 2: 'two'}, 'b': {3: 'three', 4: 'four'}}
x = d['b']
print(x[3])
```

- (●) three ✓
- ( ) {3: 'three', 4: 'four'}
- ( ) RuntimeError
- ( ) None of the above

✓ What is the output of the given code 1/1

```
import pandas as pd
import numpy as np

data = pd.DataFrame({'Name': ['nicki', 'hary', 'naman', 'jivit'],
'Age': [28, 24, 35, 32]
})
data['newcol'] = np.random.randint(1, 100, data.shape[0])

print(data['newcol'])
```

- (●) Lists random numbers assigned for each row ✓
- ( ) Lists 100 rows with value 1
- ( ) Lists all rows of newcol assigned values of 100
- ( ) RuntimeError

✗ What is the output of the following NumPy code? import numpy as np arr = np.array([1, 2, 3, 4, 5]) result = np.where(arr > 3)[0] 0/1

- ○ Array([3, 4])
- ◉ Array([4, 3]) ✗
- ○ Array([0, 1, 2])
- ○ Array([1, 2])

Correct answer

- ◉ Array([3, 4])

✗ **Which technique is used to combat the problem of class imbalance in a classification problem by assigning different misclassification costs to different classes?** 0/1

- ○ Over-sampling
- ○ Under-sampling
- ○ Cost-sensitive learning
- ◉ Bagging ✗

Correct answer

- ◉ Cost-sensitive learning

✓ What method is used for slicing a dataframe label based 1/1

- ○ iloc
- ◉ loc ✓
- ○ label
- ○ ix

✗ Given a 1D NumPy array arr, how can you find the index of the first occurrence of a value greater than a specified threshold threshold? 0/1

- ○ index = np.where(arr > threshold)[0]
- ○ index = arr.index(arr > threshold)[0]
- ○ index = np.searchsorted(arr, threshold, side='right')
- ◉ index = np.argwhere(arr > threshold)[0] ✗

Correct answer

- ◉ index = np.where(arr > threshold)[0]

✗ **Which machine learning algorithm is particularly well-suited for solving problems involving sequences and time series data, such as speech recognition or language translation?** 0/1

- ○ Decision Trees
- ○ Convolutional Neural Networks (CNN)
- ○ Long Short-Term Memory (LSTM)
- ◉ K-Means Clustering ✗

Correct answer

- ◉ Long Short-Term Memory (LSTM)

✗ Given a 2D NumPy array arr, how can you obtain the indices of the minimum value in each column? 0/1

- ○ col_min_indices = np.argmin(arr, axis=0)
- ◉ col_min_indices = arr.argmin(axis=1) ✗
- ○ col_min_indices = np.argmin(arr, axis=1)
- ○ col_min_indices = arr.argmin(axis=0)

Correct answer

- ◉ col_min_indices = np.argmin(arr, axis=0)

✓ What is the output of the given code 1/1

```
s = {10, 20, 35}
t = frozenset(s)
t.add(45)
print(t)
```

- ○ {10,20,35}
- ○ {10,20,35,45}
- ◉ TypeError ✓
- ○ None of the above

✓ What is the value of x 1/1

```
x=max("It might rain today")
```

- ◉ y ✓
- ○ might
- ○ today
- ○ it

✗ What is the output of the given code 0/1

```
a=110
if a<111:
print(‘“Negative”’)
print(a)
```

- 110
- 111
- Negative
- "Negative" ✗

Correct answer

- 110

✓ What is the output of the given code 1/1

```
def fun(num):
if(num<0):
return
if num%2==0:
print("fun")
else:
print("no fun")
num-=1
fun(num)

fun(3)
```

- ○ no fun fun
- ○ no fun fun no fun fun no fun fun
- ◉ no fun fun no fun fun ✓
- ○ RuntimeError

✓ **What is the primary goal of supervised machine learning?** 1/1

- ○ Uncover hidden patterns in data
- ○ Minimize model complexity
- ◉ Make predictions or classify data ✓
- ○ Reduce overfitting

✓ **. In a support vector machine (SVM), what is the primary goal when choosing the hyperplane that separates data points?** 1/1

- (●) Maximize the margin between data points and the hyperplane ✓
- ( ) Minimize the number of support vectors
- ( ) Minimize the dimensionality of the data
- ( ) Maximize the overlap between different classes

✗ **What is the main purpose of a validation set in the machine learning workflow?** 0/1

- ( ) To train the model
- (●) To evaluate the model's performance on unseen data ✗
- ( ) To fine-tune hyperparameters
- ( ) To test the model's generalization

Correct answer

- (●) To fine-tune hyperparameters

✓ What is the output of the given code 1/1

```
import pandas as pd
s1 = pd.Series([11,22,44,55])
s2 = s1 *2
print(s2)
```

- ○ 0 22 1 44 2 88 3 110 4 22 5 44 6 88 7 110 dtype: int64
- ◉ 0 22 1 44 2 88 3 110 dtype: int64 ✓
- ○ TypeError
- ○ None of the above

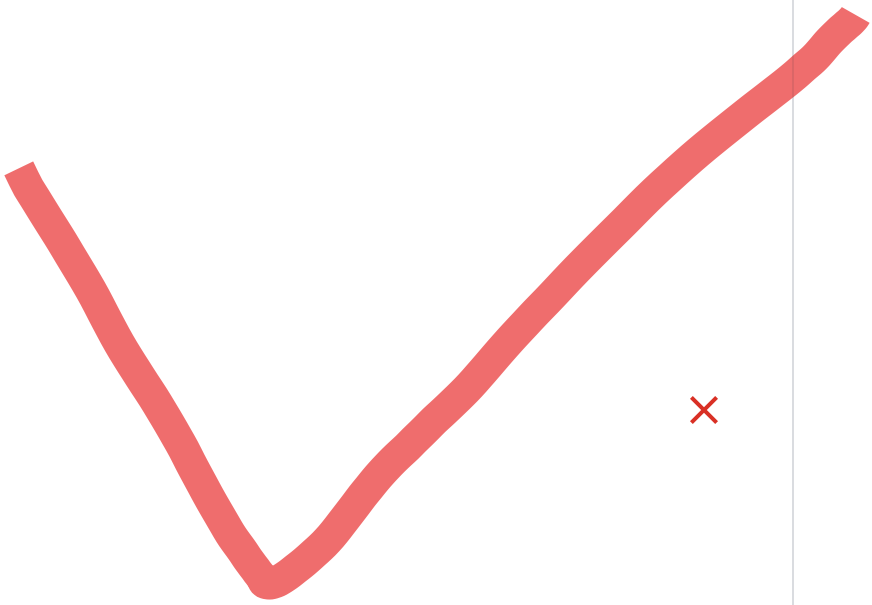
✗ What is the value of x 0/1

```
s = {1.0, "Hello", (1, 2)}
x = s[2][1]
```

- ○ 2
- ○ H
- ○ e
- ◉ TYpeError ✗

Correct answer

- ◉ e



✗ **What is the purpose of cross-validation in machine learning?** 0/1

- ○ Select the best hyperparameters for a model
- ○ Train a model on multiple datasets
- ◉ Evaluate a model's performance on a holdout dataset ✗
- ○ Assess a model's generalization and reduce bias

Correct answer

- ◉ Assess a model's generalization and reduce bias

✗ What is the output of the given code 0/1

```
d1 = {'a': 11, 'b': 22, 'c': 33, 'd': 44}
d2 = {'a': 10, 'b': 20, 'c': 30, 'd': 40}
result = map(lambda x, y: x + y, d1.values(), d2.values())
print(list(result))
```

- ◉ [11,22,33,44] ✗
- ○ [21,42,63,83]
- ○ [10,20,30,40]
- ○ none of the above

Correct answer

- ◉ [21,42,63,83]

✓ What is the value of x 1/1

d = {'a': [1, 2,5,10,20], 'b': [3, 4,1]}

x = d['a'] if 'a' in d else 'Not Found'

- (●) [1, 2,5,10,20] ✓
- ( ) TRUE
- ( ) {'a': [1, 2,5,10,20]}
- ( ) none

✗ You have a 2D NumPy array matrix. How can you calculate the mean of each column and store the result in a 1D NumPy array column_means? 0/1

- ( ) column_means = np.mean(matrix, axis=0)
- (●) column_means = matrix.mean(axis=1) ✗
- ( ) column_means = np.mean(matrix, axis=1)
- ( ) column_means = matrix.mean(axis=0)

Correct answer

- (●) column_means = np.mean(matrix, axis=0)

✓ What is the output of the given code 1/1

```
f = lambda x, y: x/10 if x > y else y/10
print(f(11, 10))
```

- (●) 1.1 ✓
- ( ) 1
- ( ) 0
- ( ) 10

✓ **Which type of machine learning algorithm is most suitable for time series forecasting tasks, such as stock price prediction or weather forecasting?** 1/1

- ( ) Supervised learning
- ( ) Unsupervised learning
- ( ) Reinforcement learning
- (●) Recurrent Neural Networks (RNN) ✓

✓ What is the correct code to list columns names of a dataframe 1/1

- ( ) df.columns.tolist()
- ( ) list(df.columns)
- (●) both a and b ✓
- ( ) None of the above

✓ **In machine learning, what does the term "hyperparameter" refer to?** 1/1

- ○ The features of the dataset
- ○ The parameters of the machine learning model
- ○ The variables that are automatically learned during training
- ◉ The settings or configurations that are set before training a model ✓

✗ **Which machine learning algorithm is best suited for anomaly detection tasks, such as fraud detection or network intrusion detection?** 0/1

- ◉ . K-Means Clustering ✗
- ○ Support Vector Machines (SVM)
- ○ Decision Trees
- ○ Naïve Bayes

Correct answer

- ◉ Support Vector Machines (SVM)

✓ **In deep learning, what is the purpose of an activation function in a neural network?** 1/1

- ○ It defines the learning rate of the network.
- ○ It initializes the weights of the network.
- ◉ It introduces non-linearity to the model. ✓
- ○ It determines the network architecture

✓ **What is the purpose of the bias-variance trade-off in model selection?** 1/1

- ○ To minimize both bias and variance
- ○ To maximize both bias and variance
- ◉ To balance the trade-off between underfitting and overfitting ✓
- ○ To ignore bias and focus only on variance