MLP

Section Id :	64065351361
Section Number :	13
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	21
Number of Questions to be attempted :	21
Section Marks :	50
Display Number Panel :	Yes
Section Negative Marks :	0
Group All Questions :	No
Enable Mark as Answered Mark for Review and Clear Response :	Yes
Maximum Instruction Time :	0
Sub-Section Number :	1
Sub-Section Id :	640653107663
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Number : 193 Question Id : 640653737474 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 0

Question Label : Multiple Choice Question

THIS IS QUESTION PAPER FOR THE SUBJECT "DIPLOMA LEVEL : MACHINE LEARNING PRACTICE (COMPUTER BASED EXAM)"

ARE YOU SURE YOU HAVE TO WRITE EXAM FOR THIS SUBJECT?

CROSS CHECK YOUR HALL TICKET TO CONFIRM THE SUBJECTS TO BE WRITTEN.

(IF IT IS NOT THE CORRECT SUBJECT, PLS CHECK THE SECTION AT THE TOP FOR THE SUBJECTS

REGISTERED BY YOU)

Options :

6406532468179. ✔ YES

6406532468180. ✖ NO

Sub-Section Number :	2
Sub-Section Id :	640653107664
Question Shuffling Allowed :	No
Is Section Default? :	null

Question Id : 640653737475 Question Type : COMPREHENSION Sub Question Shuffling Allowed : No Group Comprehension Questions : No Question Pattern Type : NonMatrix Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Question Numbers : (194 to 198)

Question Label : Comprehension

```
>>> import pandas as pd
>>> df = pd.read_csv('titanic.csv')
>>> print(df)
```

	sex	age	sibsp	parch	fare	class	embark_town	alive	alone
0	male	22.0	1	0	7.2500	Third	Southampton	no	False
1	female	38.0	1	0	71.2833	First	Cherbourg	yes	False
2	female	26.0	0	0	7.9250	Third	Southampton	yes	True
3	female	35.0	1	0	53.1000	First	Southampton	yes	False
4	male	35.0	0	0	8.0500	Third	Southampton	no	True
5	male	NaN	0	0	8.4583	Third	Queenstown	no	True
6	male	54.0	0	0	51.8625	First	Southampton	no	True
7	male	2.0	3	1	21.0750	Third	Southampton	no	False
8	female	27.0	0	2	11.1333	Third	Southampton	yes	False
9	female	14.0	1	0	30.0708	Second	Cherbourg	yes	False
10	female	4.0	1	1	16.7000	Third	Southampton	yes	False
11	female	58.0	0	0	26.5500	First	Southampton	yes	True
12	male	20.0	0	0	8.0500	Third	Southampton	no	True
13	male	39.0	1	5	31.2750	Third	Southampton	no	False
14	female	14.0	0	0	7.8542	Third	Southampton	no	True
15	female	55.0	0	0	16.0000	Second	Southampton	yes	True

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 194 Question Id : 640653737476 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Choose the correct options to determine the count of NULL values in each column of a pandas DataFrame named 'df'

Options :

6406532468181. ✔ `df.isnull().sum()`

6406532468182. ✖ `df.null().sum()`

6406532468183. ✖ `df.NA().sum()`

6406532468184. ✖ `df.isNAN().sum()`

Question Number : 195 Question Id : 640653737477 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Which option will help in selecting data of **odd indexed** columns only? (ie. age, parch are odd indexed 1 and 3 respectively)

Options :

6406532468185. ✖ `df.iloc[1::2]`

6406532468186. ✔ `df.iloc[:,1::2]`

6406532468187. ✖ `df[:,1::2]`

6406532468188. ✖ `df.loc[:,1:2]`

Question Number : 196 Question Id : 640653737478 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

What is datatype of sex column in the given titanic dataset

Options :

6406532468189. ✖ number

6406532468190. ✖ bool

6406532468191. ✖ str

6406532468192. ✔ object

Question Number : 197 Question Id : 640653737479 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Which option will help in filtering data to find **females who are alive** after titanic incident?

Options :

6406532468193. ✔ `df[(df['sex'] == 'female') & (df['alive'] == 'yes')]`

6406532468194. ✖ `df.select[((df['sex'] == 'female') & (df['alive'] == 'yes'))]`

6406532468195. ✖ `df.isin[(df['sex'] == 'female') & (df['alive'] == 'yes')]`

6406532468196. ✖ `df.group((df['sex'] == 'female') & (df['alive'] == 'yes'))`

Question Number : 198 Question Id : 640653737480 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

What is the given code below trying to accomplish for the given titanic dataset ?

```
>>> def is_adult (age):
...     if age > 18:
...         return True
...     return False
>>> df['isAdult'] = df['age'].apply(is_adult)
```

Options :

6406532468197. ✖ It will print the rows which have age greater than 18.

6406532468198. ✖ It will throw an error since while calling the function input variable is not given.

6406532468199. ✔ A new column named `'isAdult'` will be added to the dataframe which will contain True and False based on age column.

6406532468200. ✖ None of these

Sub-Section Number : 3

Sub-Section Id : 640653107665

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 199 Question Id : 640653737491 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 1

Question Label : Multiple Choice Question

What will be the output of the following code:

```
from sklearn.feature_extraction import DictVectorizer
X = [{'feature_1': 3, 'feature_2': 1}, {'feature_1': 2, 'feature_3': 3}]
extractor = DictVectorizer(sparse= False)
print(extractor.fit_transform(X))
```

Options :

6406532468245. ✖
```
[[3. 1.]
 [2. 3.]]
```

6406532468246. ✖
```
[[3. 1.]
 [3. 0.]]
```

6406532468247. ✔
```
[[3. 1. 0.]
 [2. 0. 3.]]
```

6406532468248. ✖
```
[[3. 1. 0.]
 [2. 3. 0.]]
```

Sub-Section Number :	4
Sub-Section Id :	640653107666
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 200 Question Id : 640653737481 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

You are working on a machine learning project that aims to predict housing prices based on various features of the houses. As the first step, you decide to perform exploratory data analysis and visualize the data to understand its structure and relationships. Which of the following visualization techniques or principles is LEAST likely to provide meaningful insights for this kind of

regression problem?

Options :

6406532468201. ✖ Plotting a heatmap of the correlation matrix to understand the linear relationship between the numeric features.

6406532468202. ✖ Using a scatter plot to visualize the relationship between the square footage of a house and its price.

6406532468203. ✔ Visualizing the distribution of housing prices using a pie chart.

6406532468204. ✖ Creating box plots for housing prices, grouped by the number of bedrooms, to detect outliers and understand the distribution across different categories.

Question Number : 201 Question Id : 640653737482 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

A company collects 40000 samples (examples) to build a Machine Learning model for an application. They decide to use 30% of the total samples for testing (to be stored in the variable **testset**) and the rest 70% for training (to be stored in the variable **trainset**). They also want to sample the same set of samples across multiple runs. Which of the following line (statement) achieves this task? Assume that all samples are stored in the variable **data**.

Options :

6406532468205. ✖ `testset, trainset = train_test_split(data, test_size=0.3,random_state=42)`

6406532468206. ✖ `trainset, testset = train_test_split(data, test_size=0.3)`

6406532468207. ✔ `trainset, testset = train_test_split(data, test_size=0.3,random_state=42)`

6406532468208. ✖ `testset, trainset = train_test_split(data, test_size=0.3)`

Question Number : 202 Question Id : 640653737483 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Choose the option based on the following statements:

Statement 1: The train-test split allows us to simulate the model's performance on new data by reserving a portion of the dataset for testing.

Statement 2: Separating the dataset into training and testing sets helps prevent data leakage, where information from the test set unintentionally influences the model during training. This ensures a fair assessment of the model's generalization capabilities.

Options :

6406532468209. ✖ Statement 1 is True and statement 2 is False

6406532468210. ✖ Statement 1 is False and statement 2 is True

6406532468211. ✖ Both the statements are False

6406532468212. ✔ Both the statements are True

Question Number : 203 Question Id : 640653737484 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

In which of the following scenarios is data cleaning most required?

Options :

6406532468213. ✔ Data has missing values and categorical string data.

6406532468214. ✖ Data consists a lot of outliers based on values in target(y) column

6406532468215. ✖ The data consists solely of numerical values and are on a similar scale.

6406532468216. ✖ None of the choices

Question Number : 204 Question Id : 640653737486 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Consider the below code:

```
data = [[-3, 1],
        [-3, 1],
        [ 3, 5],
        [ 3, 5]]

from sklearn.preprocessing import StandardScaler
ss = StandardScaler()
print(ss.fit_transform(data))
```

What will be the output of the code snippet given above?

Options :

6406532468221. ✖
```
[[0, -1],
 [0, -1],
 [1,  1],
 [1,  1]]
```

6406532468222. ✖
```
[[-0.5, -2],
 [-0.5, -2],
 [  1,   2],
 [  1,   2]]
```

6406532468223. ✖
```
[[ 0, 1],
 [ 0, 1],
 [ 0, 1],
 [ 0, 1]]
```

6406532468224.

```
[[-1, -1],
 [-1, -1],
 [ 1,  1],
 [ 1,  1]]
```

Question Number : 205 Question Id : 640653737487 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Which of the following metrics indicate higher their value, better the regression model's performance?

Options :

6406532468225. ✖ RMSE

6406532468226. ✔ R2

6406532468227. ✖ Mean absolute error

6406532468228. ✖ Mean squared error

Question Number : 206 Question Id : 640653737488 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

How many models with different combinations of parameter values will get trained in the following code?

```
from sklearn.model_selection import GridSearchCV
from sklearn.linear_model import SGDRegressor
from sklearn.datasets import load_diabetes

X, y = load_diabetes(return_X_y=True)

params = [
    {'alpha': [0.01,0.1,1],'learning_rate': ['constant','optimal']},
    {'loss' : ['squared_error', 'huber'], 'alpha':
    ↪  [0.0001,0.001],'learning_rate':['constant','invscaling']}]

grid= GridSearchCV(estimator= SGDRegressor(),
                   param_grid = params,
                   scoring = 'neg_mean_squared_error',
                   return_train_score=True,
                   verbose = 2,
                   n_jobs= -1
                   )

grid.fit(X,y)
```

Options :

6406532468229. ✖ 10

6406532468230. ✖ 12

6406532468231. ✔ 14

6406532468232. ✖ 16

Question Number : 207 Question Id : 640653737489 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

What is the purpose of the *tol* parameter of the `SGDRegressor()` in the given code below?

```
from sklearn.linear_model import SGDRegressor
model = SGDRegressor(early_stopping=True,
                     validation_fraction=0.2,
                     tol=0.001,
                     n_iter_no_change=5)
model.fit(X, y)
```

Options :

6406532468233. ✖ It controls the learning rate of the stochastic regressor during training.

6406532468234. ✖ It determines the maximum number of iterations for the training process.

6406532468235. ✖ It defines the fraction of the validation set used for early stopping.

6406532468236. ✔ It specifies the tolerance level for early stopping based on the change in the validation error.

Question Number : 208 Question Id : 640653737490 Question Type : MCQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Multiple Choice Question

Consider the following code:

```
from sklearn.datasets import make_regression
from sklearn.datasets import make_classification
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import LogisticRegression

X_r, y_r = make_regression()
lr = LinearRegression()
lr.fit(X_r, y_r)
score1 = lr.score(X_r, y_r)

X_c, y_c = make_classification()
logr = LogisticRegression()
logr.fit(X_c, y_c)
score2 = logr.score(X_c, y_c)

print(score1)
print(score2)
```

Which metrics will be contained in score1 and score2 respectively?

Options :

6406532468237. ✖ Accuracy, Accuracy

6406532468238. ✖ R2 score, R2 score

6406532468239. ✖ Accuracy, R2 score

6406532468240. ✔ R2 score, Accuracy

6406532468241. ✖ F1 score, Precision

6406532468242. ✖ Precision, Recall

6406532468243. ✖ MAE, MSE

6406532468244. ✖ The code will result in an error.

Sub-Section Number :	5
Sub-Section Id :	640653107667
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 209 Question Id : 640653737485 Question Type : MCQ Is Question

Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3

Question Label : Multiple Choice Question

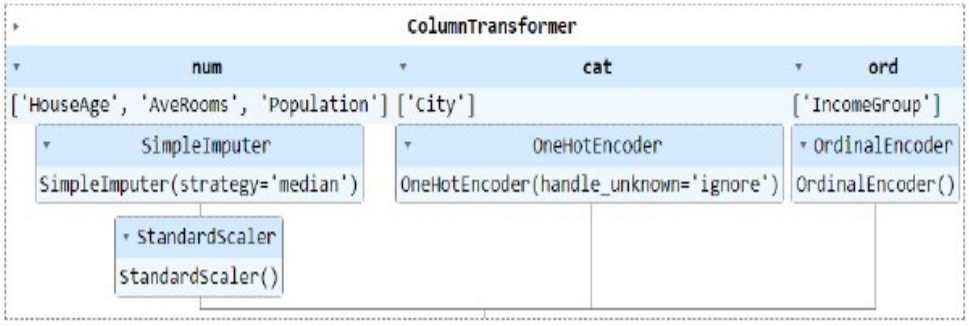
Consider following dataset:

HouseAge	AveRooms	Population	City	IncomeGroup
41.0	6.98	322.0	Delhi	Low
21.0	6.23	2401.0	Kolkata	High
52.0	8.28	496.0	Agra	Medium
52.0	-	558.0	Kolkata	Medium
52.0	6.28	565.0	Mumbai	Medium

Which one of the following code snippets will correctly preprocess above data? Assume necessary imports. The data is stored in a dataframe named X.

Options :

```
num_tranform = Pipeline(steps= [("imputer", SimpleImputer(strategy="median")),
                                ("scaler", StandardScaler())])
cat_transfom = OneHotEncoder(handle_unknown="ignore")
ordinal_encoder = OrdinalEncoder()
preprocessor = ColumnTransformer(transformers=[
                    ("num", num_tranform, ['HouseAge','AveRooms','Population']),
                    ("cat", cat_transfom, ['City']),
                    ("ord", ordinal_encoder, ['IncomeGroup'])])
preprocessor
```

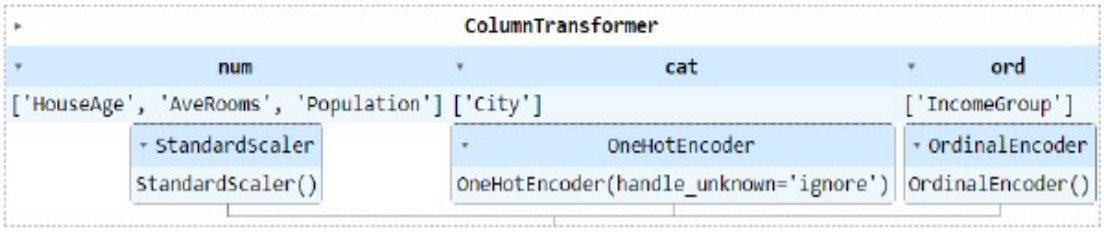


6406532468217. ✔

6406532468218. ✖

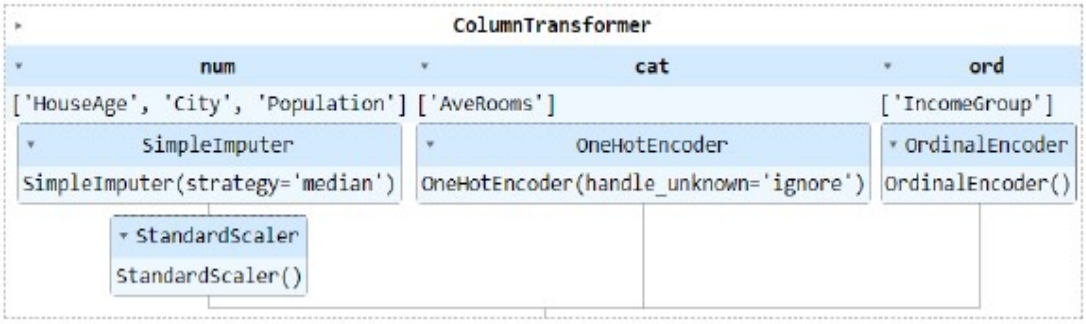
```python
num_tranform = Pipeline(steps=[("scaler", StandardScaler())])
cat_transfom = OneHotEncoder(handle_unknown="ignore")
ordinal_encoder = OrdinalEncoder()

preprocessor = ColumnTransformer(transformers=
        [("num", num_tranform, ['HouseAge','AveRooms','Population']),
         ("cat", cat_transfom, ['City']),
         ("ord", ordinal_encoder, ['IncomeGroup'])])
preprocessor
```




```python
num_tranform = Pipeline(
    steps=[("imputer", SimpleImputer(strategy="median")),
           ("scaler", StandardScaler())])

cat_transfom = OneHotEncoder(handle_unknown="ignore")
ordinal_encoder = OrdinalEncoder()
preprocessor = ColumnTransformer(transformers=
  [("num", num_tranform, ['HouseAge','City','Population']),
   ("cat", cat_transfom, ['AveRooms']),
   ("ord", ordinal_encoder, ['IncomeGroup'])])
preprocessor
```



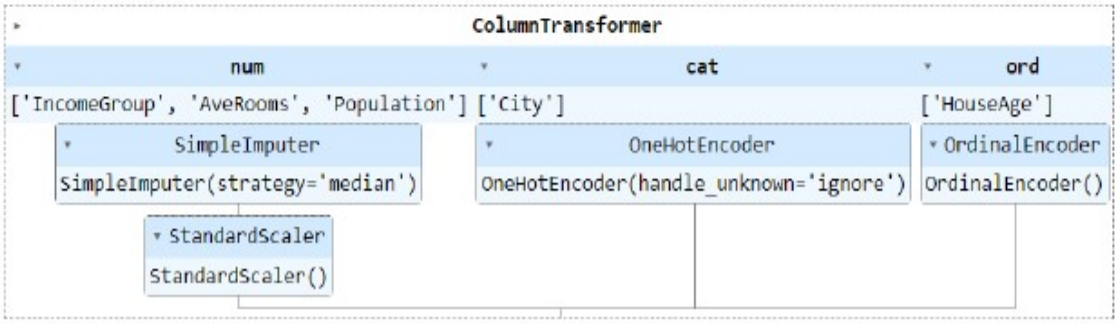
6406532468219. ✖

6406532468220. ✖

```
num_tranform = Pipeline(
    steps=[("imputer", SimpleImputer(strategy="median")),
           ("scaler", StandardScaler())]
    )

cat_transfom = OneHotEncoder(handle_unknown="ignore")
ordinal_encoder = OrdinalEncoder()

preprocessor = ColumnTransformer(transformers=
    [("num", num_tranform, ['IncomeGroup','AveRooms','Population']),
     ("cat", cat_transfom, ['City']),
     ("ord", ordinal_encoder, ['HouseAge'])
    ]
)
```



Sub-Section Number : 6

Sub-Section Id : 640653107668

Question Shuffling Allowed : Yes

Is Section Default? : null

Question Number : 210 Question Id : 640653737492 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

Suppose that we plot the histogram of numerical features in a data set. This reveals which of the following information?

Options :

6406532468249. ✔ Scale of the features

6406532468250. ✔ (left or right) Skew of the distribution

6406532468251. ✔ Modes in the distribution

6406532468252. ✖ Missing values

Question Number : 211 Question Id : 640653737493 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

Choose the correct option(s) for the below code

```
import numpy as np
from sklearn.datasets import fetch_california_housing
house = fetch_california_housing(as_frame=True)
```

Options :

6406532468253. ✖ **house.data.shape** will give the count of number of rows and columns for features and labels(target) both

6406532468254. ✔ **house.target_names** will give the name of the target column

6406532468255. ✖ **house.data.head()** will show last 5 rows of the data

6406532468256. ✔ **house.feature_names** will give a list of names of the columns of the feature matrix

Question Number : 212 Question Id : 640653737494 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

Which of the following are valid loss functions for SGDClassifier?

Options :

6406532468257. ✖ Squared Loss

6406532468258. ✔ Hinge Loss

6406532468259. ✔ Log Loss

6406532468260. ✖ Mean Absolute Error

Question Number : 213 Question Id : 640653737495 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

Which columns will not be included in the selected data within the code below?

```
import numpy as np
from sklearn.feature_selection import VarianceThreshold

data = np.array([[  95,  0.332,   112,    1,    0.56  ],
       [ 146,  0.332,   177,    1,    9.2   ],
       [ -96,  0.332,  -139,    1,   -0.82  ],
       [ 116,  0.332,   117,    1,    4.8   ],
       [ -87,  0.332,   -63,    1,   -1.1   ]])

vf = VarianceThreshold(threshold=0)

selected_data = vf.fit_transform(data)
selected_data
```

Options :

6406532468261. ✖ Column indexed at 0

6406532468262. ✔ Column indexed at 1

6406532468263. ✖ Column indexed at 2

6406532468264. ✔ Column indexed at 3

6406532468265. ✖ Column indexed at 4

6406532468266. ✖ No columns

Question Number : 214 Question Id : 640653737498 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

Which of the following Evaluation metrics can be used in a regression problem?

Options :

6406532468277. ✔ Mean Squared Error

6406532468278. ✖ Accuracy

6406532468279. ✖ F1-Score

6406532468280. ✔ Mean Absolute Error

6406532468281. ✖ credit score

6406532468282. ✖ CGPA

Sub-Section Number :	7
Sub-Section Id :	640653107669
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 215 Question Id : 640653737496 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3 Max. Selectable Options : 0

Question Label : Multiple Select Question

Consider the below code :

keep following symbols in mind:

- >>>: Represents input code
- # : Represents comment in a code
- ... : Represents code continuation
- Without any symbols at the beginning of a line then it is output of just above input line of code
- arrow pointing towards right is code continuation to another line.

```
>>> from sklearn.feature_selection import SelectKBest, chi2
>>> from sklearn.datasets import load_wine
>>> X,y = load_wine(return_X_y=True,as_frame=True)
>>> print(X.shape)
(178, 13)

>>> print(X.columns)
['alcohol', 'malic_acid', 'ash', 'alcalinity_of_ash', 'magnesium',
 'total_phenols', 'flavanoids', 'nonflavanoid_phenols', 'proanthocyanins',
 'color_intensity', 'hue', 'od280/od315_of_diluted_wines', 'proline']

>>> skb = SelectKBest(chi2, k=3)
>>> X_selected = skb.fit_transform(X, y)

>>> print(skb.scores_)
[5.44, 28.06, 0.74, 29.38, 45.02, 15.62, 63.33, 1.81,
 9.36, 109.01, 5.18, 23.38, 16540.06]

>>> print(skb.pvalues_)
[6.56e-02, 8.03e-07, 0.68, 4.16e-07, 1.66e-10, 4.05e-04,
 1.76e-14,  0.40, 9.24e-03, 2.12e-24, 0.074, 8.33e-06, 0]

>>> print(skb.pvalues_.argsort())
[12,  9,  6,  4,  3,  1, 11,  5,  8,  0, 10,  7,  2]
```

Which of the following feature(s) will be selected in X_selected from X ?

Options :

6406532468267. ✖ malic_acid

6406532468268. ✖ magnesium

6406532468269. ✔ flavanoids

6406532468270. ✖ proanthocyanins

6406532468271. ✔ color_intensity

6406532468272. ✔ proline

Question Number : 216 Question Id : 640653737497 Question Type : MSQ Is Question Mandatory : No Calculator : None Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 3 Max. Selectable Options : 0

Question Label : Multiple Select Question

Given the following code snippet, which statement is true regarding the use of LabelEncoder?

```
from sklearn.preprocessing import LabelEncoder

data = ["cat", "dog", "fish", "cat", "bird", "dog", "bird"]

encoder = LabelEncoder()
encoded_data = encoder.fit_transform(data)

print(encoded_data)
```

Options :

6406532468273. ✔ encoded_data will contain only the values 0, 1, 2, and 3.

6406532468274. ✖ If ["elephant"] is passed to encoder.transform(), it will be successfully transformed to an integer.

6406532468275. ✖ LabelEncoder assigns higher integer values to more frequently occurring labels.

6406532468276. ✔ After calling encoder.inverse_transform([2]), the result will be "dog".

Sub-Section Number :	8
Sub-Section Id :	640653107670
Question Shuffling Allowed :	Yes
Is Section Default? :	null

Question Number : 217 Question Id : 640653737499 Question Type : SA Calculator : None

Response Time : N.A Think Time : N.A Minimum Instruction Time : 0

Correct Marks : 2

Question Label : Short Answer Question

What will be the output of the following code ?

```
data = [['apple', 120],
        ['apple', 125],
        ['grapes',120]]

from sklearn.preprocessing import OneHotEncoder
ohe = OneHotEncoder(sparse_output=False)
ohe.fit(data)
print(ohe.transform(data).shape[1])
```

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

4

Business Analytics

Section Id :	64065351362
Section Number :	14
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	10
Number of Questions to be attempted :	10
Section Marks :	20
Display Number Panel :	Yes