6406532775036. ✖ $\prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$

6406532775037. ✖ $\sum_{i=1}^{n} [-\lambda + \lambda \log x_i - \log(x_i!)]$

6406532775038. ✖ $\prod_{i=1}^{n} [-\lambda + x_i \log \lambda - \log(x_i!)]$

Question Number : 132 Question Id : 640653825126 Question Type : SA

Correct Marks : 2.5

Question Label : Short Answer Question

Consider a dataset that has 25 data-points. The data-point x_i and its frequency is given in the following table:

x_i	Frequency
0	1
1	4
2	6
3	9
4	5

In case the table is not clear: the value 0 appears once in the dataset, the value 1 appears four times in the dataset, and so on. Find the maximum likelihood estimate for the parameter λ of the Poisson distribution given this dataset.

Response Type : Numeric

Evaluation Required For SA : Yes

Show Word Count : Yes

Answers Type : Equal

Text Areas : PlainText

Possible Answers :

2.52

MLP

Section Id :	64065359216
Section Number :	9
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	14
Number of Questions to be attempted :	14
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 :	No
Section Maximum Duration :	0
Section Minimum Duration :	0
Section Time In :	Minutes
Maximum Instruction Time :	0
Sub-Section Number :	1
Sub-Section Id :	640653122758
Question Shuffling Allowed :	No

Question Number : 133 Question Id : 640653825130 Question Type : MCQ

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 :

6406532775045. ✔ YES

6406532775046. ✖ NO

Sub-Section Number :	2
Sub-Section Id :	640653122759
Question Shuffling Allowed :	No

Question Id : 640653825131 Question Type : COMPREHENSION Sub Question Shuffling Allowed : No Group Comprehension Questions : No Question Pattern Type : NonMatrix

Question Numbers : (134 to 139)

Question Label : Comprehension

Consider following common data and answer the given subquestions:

```
import pandas as pd
import numpy as np
columns = ["Name", "Manufacturer", "Year_of_purchase",
           "Price_in_lacs","Fuel_type"]
data = [["800 AC", "Maruti", 2007, 0.6, "Petrol", ],
        ["Wagon R", "Maruti", 2007, 1.35, "Petrol", ],
        ["Verna", "Hyundai", 2012, 6.0, "Diesel", ],
        ["Corolla", "Toyota", 2018, 16.5, "Petrol", ],
        ["Amaze", "Honda", 2014, 4.5, "Diesel", ],
        ["Alto", "Maruti", 2007, 1.4, "Petrol", ],]
df = pd.DataFrame(data=data, columns=columns)
```

Sub questions

Question Number : 134 Question Id : 640653825132 Question Type : SA
Correct Marks : 1
Question Label : Short Answer Question
What will be the output of the following code snippet?

```
print(np.sum(df.shape))
```

Enter -1, if you think the above code snippet will generate an error.
Response Type : Numeric
Evaluation Required For SA : Yes
Show Word Count : Yes
Answers Type : Equal
Text Areas : PlainText
Possible Answers :
11

Question Number : 135 Question Id : 640653825133 Question Type : SA
Correct Marks : 1
Question Label : Short Answer Question
What will be the output of the following code snippet?

```
t = df['Year_of_purchase'].value_counts()
print(t[2007])
```

Enter -1, if you think the above code snippet will generate an error.
Response Type : Numeric
Evaluation Required For SA : Yes
Show Word Count : Yes
Answers Type : Equal

Text Areas : PlainText
Possible Answers :
3

Question Number : 136 Question Id : 640653825134 Question Type : SA
Correct Marks : 2
Question Label : Short Answer Question
What will be the output of the following code snippet?

```
print(df.describe().T.shape[0])
```

Enter -1, if you think the above code snippet will generate an error.
Response Type : Numeric
Evaluation Required For SA : Yes
Show Word Count : Yes
Answers Type : Equal
Text Areas : PlainText
Possible Answers :
2

Question Number : 137 Question Id : 640653825135 Question Type : SA
Correct Marks : 2
Question Label : Short Answer Question
What will be the output of the following code snippet?

```
t=df.query("Year_of_purchase >=2008 and Fuel_type!='Petrol'")
print(t.shape[0])
```

Enter -1, if you think the above code snippet will generate an error.
Response Type : Numeric
Evaluation Required For SA : Yes
Show Word Count : Yes
Answers Type : Equal
Text Areas : PlainText
Possible Answers :
2

Question Number : 138 Question Id : 640653825136 Question Type : MSQ
Correct Marks : 2 Max. Selectable Options : 0
Question Label : Multiple Select Question
Given the data, which of the following options will provide the same output?

Options :

6406532775051. ✔ `df.iloc[3]['Year_of_purchase']`

6406532775052. ✔ `df.loc[3,'Year_of_purchase']`

6406532775053. ✔ `df.iloc[3,2]`

6406532775054. ✔ `df.iloc[-3,2]`

6406532775055. ✔ `df[df['Price_in_lacs'] >=10.0].Year_of_purchase.iloc[0]`

6406532775056. ✖ All options will provide different outputs.

Question Number : 139 Question Id : 640653825137 Question Type : MSQ

Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

If you would like to see the most frequent car manufacturers in the dataset, which of the following can be used?

Options :

6406532775057. ✔ `df.Manufacturer.value_counts()`

6406532775058. ✔ `df.groupby(by='Manufacturer')['Fuel_type'].count()`

6406532775059. ✖ `df.groupby(by='Fuel_type')['Manufacturer'].count()`

6406532775060. ✖ `df.Manufacturer.count_values()`

6406532775061. ✖ None of these

Sub-Section Number : 3

Sub-Section Id : 640653122760

Question Shuffling Allowed : Yes

Question Number : 140 Question Id : 640653825138 Question Type : MSQ

Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

Which of the following choice(s) are correct ?

Options :

6406532775062. ✔ Null values cannot be interpreted by the model hence we need to handle them accordingly.

6406532775063. ✖ Null values cannot be replaced because we can not manipulate the dataset.

6406532775064. ✖ We should let the sklearn or software automatically decide how to handle different kinds of missing values.

6406532775065. ✔ Different types of representation of missing values could be seen in the dataset.

Question Number : 141 Question Id : 640653825140 Question Type : MSQ
Correct Marks : 2 Max. Selectable Options : 0
Question Label : Multiple Select Question
Which of the following options are true about Pearson correlation matrix ?
Options :

6406532775072. ✔ Correlation coefficient values in the matrix are in the range of -1 to +1.

6406532775073. ✖ Higher correlation coefficient among the features leads to better model predictions.

6406532775074. ✔ Correlation matrix can be represented graphically using the heatmap.

6406532775075. ✖ Correlation coefficient values in the matrix could be in the range of $-\infty$ to $+\infty$.

Sub-Section Number : 4
Sub-Section Id : 640653122761
Question Shuffling Allowed : Yes

Question Number : 142 Question Id : 640653825139 Question Type : MSQ
Correct Marks : 3 Max. Selectable Options : 0
Question Label : Multiple Select Question
which of the following methods come under sklearn's `model_selection` module ?
Options :

6406532775066. ✖ ColumnTransformer

6406532775067. ✔ GridSearchCV

6406532775068. ✖ Pipeline

6406532775069. ✔ StratifiedShuffleSplit

6406532775070. ✖ `mean_absolute_error`

6406532775071. ✖ trees

Sub-Section Number : 5
Sub-Section Id : 640653122762
Question Shuffling Allowed : Yes

Question Number : 143 Question Id : 640653825141 Question Type : MCQ
Correct Marks : 2
Question Label : Multiple Choice Question
We need to preprocess the data before using it for model building due to which of the following reasons ?
Options :
6406532775076. ✖ Due to errors in data capture, data may contain outliers or missing values.
6406532775077. ✖ Different features may be at different scales.
6406532775078. ✖ Data contains non numerical features.
6406532775079. ✔ All of these

Sub-Section Number : 6
Sub-Section Id : 640653122763
Question Shuffling Allowed : No

Question Id : 640653825142 Question Type : COMPREHENSION Sub Question Shuffling Allowed : No Group Comprehension Questions : No Question Pattern Type : NonMatrix
Question Numbers : (144 to 145)
Question Label : Comprehension
Consider following data and answer the given subquestions:

```
data = np.array([[1,2,3,4],
                 [3, np.nan, 4, 1],
                 [np.nan, 2,3, 5],
                 [np.nan, np.nan, np.nan, 10]])
```

Each row represents a data point. There are exactly 4 points. The index of points starts from 0 and ends at 3 (included).
Sub questions

Question Number : 144 Question Id : 640653825143 Question Type : MCQ
Correct Marks : 3
Question Label : Multiple Choice Question
Which of the following pairs have the highest euclidean distance? Note: take care of the missing values and adjust accordingly. The options refer to the indices of points.
Options :
6406532775080. ✖ 1 and 2

6406532775081. ✖ 0 and 2

6406532775082. ✖ 1 and 3

6406532775083. ✖ 2 and 3

6406532775084. ✔ 0 and 1

6406532775085. ✖ There is a tie between two or more options.

6406532775086. ✖ None of these.

Question Number : 145 Question Id : 640653825144 Question Type : MCQ

Correct Marks : 2

Question Label : Multiple Choice Question

Which of the following pairs have the smallest euclidean distance? Note: take care of the missing values and adjust accordingly. The options refer to the indices of points.

Options :

6406532775087. ✖ 1 and 2

6406532775088. ✔ 0 and 2

6406532775089. ✖ 1 and 3

6406532775090. ✖ 2 and 3

6406532775091. ✖ 0 and 1

6406532775092. ✖ There is a tie between two or more options.

6406532775093. ✖ None of these.

Sub-Section Number :	7
Sub-Section Id :	640653122764
Question Shuffling Allowed :	No

Question Id : 640653825145 Question Type : COMPREHENSION Sub Question Shuffling Allowed : No Group Comprehension Questions : No Question Pattern Type : NonMatrix

Question Numbers : (146 to 150)

Question Label : Comprehension

Consider following data and code snippet:

```
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import MinMaxScaler, OrdinalEncoder, MaxAbsScaler
from sklearn.impute import SimpleImputer
from sklearn.pipeline import Pipeline

X = np.array([[5, 4, 'cat'],
              [4, -1, 'dog'],
              [3, 1, 'bird'],
              [np.nan, 3, 'cat']])

impute_scale_pipe = Pipeline([('impute', SimpleImputer()),
                              ('scale', MinMaxScaler())])

ct = ColumnTransformer([('impute_scale', impute_scale_pipe, [0,1]),
                        ('scale_only', MaxAbsScaler(), [1]),
                        ('categorical', OrdinalEncoder(), [2])])
transformed_X = ct.fit_transform(X)
```

Based on the above data, answer the given subquestions.

Sub questions

Question Number : 146 Question Id : 640653825146 Question Type : SA
Correct Marks : 2

Question Label : Short Answer Question

What will be the output of the following code snippet:

```
print(transformed_X.shape[1])
```

Enter -1, if you think the above code snippet will generate an error.

Response Type : Numeric
Evaluation Required For SA : Yes
Show Word Count : Yes
Answers Type : Equal
Text Areas : PlainText
Possible Answers :

4

Question Number : 147 Question Id : 640653825147 Question Type : SA
Correct Marks : 2

Question Label : Short Answer Question

What will be the output of the following code snippet:

```
print(transformed_X[3,0])
```

Enter -1, if you think the above code snippet will generate an error.

Response Type : Numeric
Evaluation Required For SA : Yes
Show Word Count : Yes
Answers Type : Equal
Text Areas : PlainText
Possible Answers :
0.5

Question Number : 148 Question Id : 640653825148 Question Type : SA
Correct Marks : 2
Question Label : Short Answer Question
What will be the output of the following code snippet:

```
print(transformed_X[0,-1])
```

Enter -1, if you think the above code snippet will generate an error.
Response Type : Numeric
Evaluation Required For SA : Yes
Show Word Count : Yes
Answers Type : Equal
Text Areas : PlainText
Possible Answers :
1.0

Question Number : 149 Question Id : 640653825149 Question Type : SA
Correct Marks : 2
Question Label : Short Answer Question
What will be the output of the following code snippet:

```
print(transformed_X[2,1])
```

Enter -1, if you think the above code snippet will generate an error.
Response Type : Numeric
Evaluation Required For SA : Yes
Show Word Count : Yes
Answers Type : Equal
Text Areas : PlainText
Possible Answers :
0.4

Question Number : 150 Question Id : 640653825150 Question Type : MSQ

Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

Which of the following can be used to get the simpleimputer object?

Options :

6406532775098. ✖
```
impute_scale_pipe.steps[1][1]
```

6406532775099. ✔
```
impute_scale_pipe['impute']
```

6406532775100. ✔
```
impute_scale_pipe[0]
```

6406532775101. ✖
```
impute_scale_pipe.get_step('impute')
```

6406532775102. ✖
```
impute_scale_pipe.get_step(0)
```

6406532775103. ✖
```
ct[0][0][0]
```

6406532775104. ✖ None of these.

Sub-Section Number :	8
Sub-Section Id :	640653122765
Question Shuffling Allowed :	No

Question Id : 640653825151 Question Type : COMPREHENSION Sub Question Shuffling Allowed : No Group Comprehension Questions : No Question Pattern Type : NonMatrix Question Numbers : (151 to 152)

Question Label : Comprehension

Go through the code snippet given below and answer the given subquestions.

```
import numpy as np
from sklearn.linear_model import SGDRegressor
from sklearn.pipeline import make_pipeline
n_samples, n_features = 18, 4
rng = np.random.RandomState(0)
y = rng.randn(n_samples)
X = rng.randn(n_samples, n_features)
reg = SGDRegressor(max_iter=1000,
                   tol=1e-3,
                   eta0= 0.04,
                   power_t=5,
                   n_iter_no_change=3,
                   validation_fraction=0.3 ,
                   random_state=42)
reg.fit(X, y)
print(reg.coef_)
```

Sub questions

Question Number : 151 Question Id : 640653825152 Question Type : MCQ
Correct Marks : 2
Question Label : Multiple Choice Question
Which of the following options will be the output of the given code?
Options :

6406532775105. ✔ [-0.02634908 0.01189399 0.0917284 0.08966849]

6406532775106. ✖ array([-0.22622766, -0.00582008, -0.1820344 , 0.03518086, -0.14490955])

6406532775107. ✖ array([-0.22622766, -0.00582008, -0.1820344])

6406532775108. ✖ Given code will return an error because the data set is not given.

Question Number : 152 Question Id : 640653825153 Question Type : MCQ
Correct Marks : 2
Question Label : Multiple Choice Question
Which of the following could be the possible output of `print(reg.score())`?
Options :

6406532775109. ✖ -0.528

6406532775110. ✖ 1

6406532775111. ✖ 0.528

6406532775112. ✔ Given code will return an error

Sub-Section Number : 9
Sub-Section Id : 640653122766
Question Shuffling Allowed : Yes

Question Number : 153 Question Id : 640653825154 Question Type : MCQ
Correct Marks : 3
Question Label : Multiple Choice Question
Consider the following code:

```
import numpy as np
from sklearn.model_selection import ShuffleSplit
data = np.random.randn(10,3) # 10x3 matrix
rs = ShuffleSplit(n_splits=3, test_size= 0.3)
for data_train, data_test in rs.split(data):
  print(data_train, data_test)
```

Which of the following may be the correct output of the above code?:
Options :

6406532775113. ✔
```
[2 9 1 7 4 5 8 0] [3 6]
[3 9 1 6 0 8 2 4] [7 5]
[8 4 7 1 6 9 0 3] [2 5]
```

6406532775114. ✖
```
[8 6 9 7 0 4 5] [3 2 1]
[1 7 6 2 5 0 3] [8 4 9]
[9 1 5 4 3 6 8] [0 7 2]
```

6406532775115. ✖
```
[2 3 7 1 0 8] [5 6 9 4]
[0 2 1 5 8 4] [3 7 6 9]
[7 6 3 9 4 0] [5 1 8 2]
```

6406532775116. ✖
```
[6 3 9 2 8] [0 4 1 5 7]
[6 7 8 4 0] [1 5 3 2 9]
[0 4 6 3 2] [9 1 7 8 5]
```

6406532775117. ✖ None of these

Question Number : 154 Question Id : 640653825156 Question Type : MCQ
Correct Marks : 3
Question Label : Multiple Choice Question
Consider the following code:

```
import numpy as np
from sklearn.linear_model import LinearRegression
X = np.array([[1, 1], [1, 2], [2, 2], [2, 3], [2, 1], [3, 3]])
# y = 1 * x_0 + 2 * x_1 + 3
y = np.dot(X, np.array([1, 2])) + 3

reg1 = LinearRegression(fit_intercept = False).fit(X, y)
s1 = reg1.score(X, y)

reg2 = LinearRegression(fit_intercept = True).fit(X, y)
s2 = reg2.score(X, y)
```

Which of the following is more likely to be true?

Options :

6406532775119. ✖ $s1 = s2$

6406532775120. ✔ $s1 < s2$

6406532775121. ✖ $s1 > s2$

Sub-Section Number :	10
Sub-Section Id :	640653122767
Question Shuffling Allowed :	Yes

Question Number : 155 Question Id : 640653825155 Question Type : SA
Correct Marks : 2

Question Label : Short Answer Question

What will be the output of the following code ?

```
from sklearn.metrics import r2_score
y_test = [4, -1, 3, 6]
y_pred = [3.5, -0.5, 2, 8]
r2_score(y_test, y_pred)
```

Response Type : Numeric
Evaluation Required For SA : Yes
Show Word Count : Yes
Answers Type : Range
Text Areas : PlainText
Possible Answers :

0.76 to 0.80

Sub-Section Number :	11
Sub-Section Id :	640653122768
Question Shuffling Allowed :	Yes

Question Number : 156 Question Id : 640653825157 Question Type : MCQ

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 :

6406532775122. ✖ It controls the learning rate of the stochastic regressor during training.

6406532775123. ✖ It determines the maximum number of iterations for the training process.

6406532775124. ✖ It defines the fraction of the validation set used for early stopping.

6406532775125. ✔ It specifies the tolerance level for early stopping based on the change in the validation error.

Sub-Section Number : 12

Sub-Section Id : 640653122769

Question Shuffling Allowed : Yes

Question Number : 157 Question Id : 640653825158 Question Type : MSQ

Correct Marks : 2 Max. Selectable Options : 0

Question Label : Multiple Select Question

Consider the following code block and if needed make appropriate assumptions:

```
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import ShuffleSplit
shuffle_split = ShuffleSplit(n_splits=5, test_size=0.2, random_state=42)
score = cross_val_score(estimator= LinearRegression(),
                        X= X_train,
                        y= y_train,
                        cv= shuffle_split,
                        scoring= '____________')
```

Which of the following may be appropriate to be filled in the blank space value for scoring parameter?

Options :

6406532775126. ✖ mean_squared_error

6406532775127. ✔ neg_mean_squared_error

6406532775128. ✔ r2

6406532775129. ✖ neg_r2

6406532775130. ✖ accuracy

6406532775131. ✖ neg_accuracy

System Commands

Section Id :	64065359217
Section Number :	10
Section type :	Online
Mandatory or Optional :	Mandatory
Number of Questions :	16
Number of Questions to be attempted :	16
Section Marks :	100
Display Number Panel :	Yes
Section Negative Marks :	0
Group All Questions :	No
Enable Mark as Answered Mark for Review and Clear Response :	No
Section Maximum Duration :	0
Section Minimum Duration :	0
Section Time In :	Minutes
Maximum Instruction Time :	0
Sub-Section Number :	1
Sub-Section Id :	640653122770
Question Shuffling Allowed :	No

Question Number : 158 Question Id : 640653825159 Question Type : MCQ

Correct Marks : 0

Question Label : Multiple Choice Question

THIS IS QUESTION PAPER FOR THE SUBJECT "DIPLOMA LEVEL : SYSTEM COMMANDS (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.