Day 1 Recap

OVERVIEW OF MACHINE LEARNING

edquest

Q1: Fill in the blank to correctly import the Linear Regression model in Python:

```
import pandas as pd
from sklearn.linear_model import ____________________
```

Choose the correct option

A: LinearRegression

B: linear_regression

C: regression

D: lin_reg

edquest

Q1: Fill in the blank to correctly import the Linear Regression model in Python:

```
import pandas as pd
from sklearn.linear_model import ____________________
```

Choose the correct option

A: LinearRegression

B: linear_regression

C: regression

D: lin_reg

edquest

Q2: Fill in the blank to correctly import the Logistic Regression model in Python:

```
import pandas as pd
from sklearn.linear_model import ____________________
```

Choose the correct option

A: Regression

B: logistic_regression

C: LogisticRegression

D: log_reg

edquest

Q2: Fill in the blank to correctly import the Logistic Regression model in Python:

```
import pandas as pd
from sklearn.linear_model import ____________________
```

Choose the correct option

A: Regression

B: logistic_regression

C: LogisticRegression

D: log_reg

edquest

Q3:Which module should you import train_test_split from in Python?

```
import pandas as pd
from _____________________________ import train_test_split
```

Choose the correct option

A. sklearn.preprocessing

B. sklearn.model_selection

C. sklearn.metrics

D. sklearn.linear_model

edquest

Q3:Which module should you import train_test_split from in Python?

```
import pandas as pd
from _____________________________ import train_test_split
```

Choose the correct option

A. sklearn.preprocessing

B. sklearn.model_selection

C. sklearn.metrics

D. sklearn.linear_model

edquest

Q4: Which of the following is the correct function used to load a .csv file into a pandas DataFrame in Python?

```
import pandas as pd
df = pd._____("ice_cream.csv")
```

Choose the correct option

A. pandas.load_csv('filename.csv')

B. pandas.open_csv('filename.csv')

C. pandas.import_csv('filename.csv')

D. pd.read_csv('filename.csv')

edquest

Q4: Which of the following is the correct function used to load a .csv file into a pandas DataFrame in Python?

```
import pandas as pd
df = pd._____("ice_cream.csv")
```

Choose the correct option

A. pandas.load_csv('filename.csv')

B. pandas.open_csv('filename.csv')

C. pandas.import_csv('filename.csv')

D. pd.read_csv('filename.csv')

edquest

Q5: What does the following line do in Python?

```
joblib.dump(model_lin_reg, 'linear_regression_model.pkl')
```

Choose the correct option

A. It saves the model to a file

B. It loads a model from a file

C. It evaluates the model and prints results

D. It deletes the model from memory

edquest

Q5: What does the following line do in Python?

```
joblib.dump(model_lin_reg, 'linear_regression_model.pkl')
```

Choose the correct option

A. It saves the model to a file

B. It loads a model from a file

C. It evaluates the model and prints results

D. It deletes the model from memory

edquest

Q6: What does test_size=0.2 indicate in the train_test_split() function?

```
X_train,X_test,Y_train,Y_test = train_test_split(X,Y,test_size=0.2,random_state=1)
```

Choose the correct option

A. The data will not be split

B. 20% of the data will be used for testing

C. 80% of the data will be used for testing

D. 20% of the data will be used for training

edquest

Q6: What does test_size=0.2 indicate in the train_test_split() function?

```
X_train,X_test,Y_train,Y_test = train_test_split(X,Y,test_size=0.2,random_state=1)
```

Choose the correct option

A. The data will not be split

B. 20% of the data will be used for testing

C. 80% of the data will be used for testing

D. 20% of the data will be used for training

edquest

Knowledge Test Day 2

EXPLORATORY DATA ANALYSIS (EDA)

edquest

1. What is the main goal of Exploratory Data Analysis (EDA)?

a) To train a model

b) To visualize charts only

c) To explore, clean, and understand the data

d) To export data to Excel

edquest

1. What is the main goal of Exploratory Data Analysis (EDA)?

a) To train a model

b) To visualize charts only

c) To explore, clean, and understand the data

d) To export data to Excel

edquest

2. Which Pandas function provides column data types and non-null counts?

a) df.describe()

b) df.shape()

c) df.info()

d) df.columns()

edquest

2. Which Pandas function provides column data types and non-null counts?

a) df.describe()

b) df.shape()

c) df.info()

d) df.columns()

edquest

3. Which Seaborn function is used to plot the distribution of a numeric column?

a) sns.countplot()

b) sns.histplot()

c) sns.barplot()

d) sns.boxplot()

edquest

3. Which Seaborn function is used to plot the distribution of a numeric column?

a) sns.countplot()

b) sns.histplot()

c) sns.barplot()

d) sns.boxplot()

edquest

4: What is the primary goal of Level 0 in Exploratory Data Analysis (EDA)?

A. To apply machine learning models

B. To clean missing and incorrect data

C. To understand the structure and content of raw data

D. To visualize the transformed data

edquest

4: What is the primary goal of Level 0 in Exploratory Data Analysis (EDA)?

A. To apply machine learning models

B. To clean missing and incorrect data

C. To understand the structure and content of raw data

D. To visualize the transformed data

edquest

5. What is feature engineering?

a) Scaling data

b) Adding noise

c) Cleaning nulls

d) Creating new meaningful features from existing data

edquest

5. What is feature engineering?

a) Scaling data

b) Adding noise

c) Cleaning nulls

d) Creating new meaningful features from existing data

edquest

6. Which type of data is used in supervised learning?

a) Labeled data

b) Unlabeled data

c) Raw image data

d) Audio data

edquest

6. Which type of data is used in supervised learning?

a) Labeled data

b) Unlabeled data

c) Raw image data

d) Audio data

edquest

7. You have a column purchase_date in string format. What is the best feature engineering step?

a) Drop the column

b) Encode it using LabelEncoder

c) Extract year, month, and day to create new features

d) Apply scaling

edquest

7. You have a column purchase_date in string format. What is the best feature engineering step?

a) Drop the column

b) Encode it using LabelEncoder

c) Extract year, month, and day to create new features

d) Apply scaling

edquest

8. In a messy dataset, which is the best FIRST step?

a) Model the data

b) Visualize correlation

c) One-hot encode all column

d) Understand structure with .info()

edquest

8. In a messy dataset, which is the best FIRST step?

a) Model the data

b) Visualize correlation

c) One-hot encode all column

d) Understand structure with .info()

edquest

9. What is a key difference between feature engineering and feature transformation?

a) Engineering is model-based, transformation is manual

b) Engineering creates new features; transformation modifies existing ones

c) Both are the same process

d) Transformation always follows feature engineering

edquest

9. What is a key difference between feature engineering and feature transformation?

a) Engineering is model-based, transformation is manual

b) Engineering creates new features; transformation modifies existing ones

c) Both are the same process

d) Transformation always follows feature engineering