Veekshith.M

PES2UG23EC807

Python For Data Science

Assignment-1

1) Given two variables, j = 6 and g = 3.3. If both normal division and floor division operators were used to divide j by g, what would be the data type of the value obtained from the operations? 1mark

- int, int
- float, float
- float, int
- int, float

Answer= float, float

2) Let a = 5 (101 in binary) and b = 3 (011 in binary). What is the result of the following operation? 1marks

```
a = 5
b = 3
p r i n t ( a & b )
```

- 3
- 7
- 5
- 1

Answer= 1

1) Explain the difference between supervised and unsupervised learning in Data Science. Give one example of Python libraries used for each?

Answer:

- **Supervised Learning**: In supervised learning, the model is trained using labeled data (input and corresponding output). Example: Predicting house prices based on size.

Python library: scikit-learn (e.g., Linear Regression).

- **Unsupervised Learning**: In unsupervised learning, the model works on unlabeled data and finds hidden patterns or clusters. Example: Customer segmentation.

Python library: scikit-learn (e.g., K-Means).
Thus, supervised → prediction with labels, unsupervised → pattern discovery without labels.

2) Describe the different steps involved in a typical Data Science workflow using Python?

Answer:

A standard workflow has the following steps:

- 1. Data Collection – Gathering datasets from CSV, databases, APIs, or sensors.
- Example: pd.read_csv("data.csv")
- 2. Data Cleaning – Handling missing values, duplicates, and errors.
- Example: df.dropna() or df.fillna(0)

- 3. Data Exploration & Visualization – Understanding the dataset using summary statistics and plots.
- 4. Feature Engineering – Selecting or creating relevant features.
- 5. Model Building – Applying ML algorithms using libraries like scikit-learn.
- 6. Model Evaluation – Checking performance using accuracy, RMSE, or confusion matrix.