Status	Finished
Started	Sunday, 2 March 2025, 7:43 PM
Completed	Sunday, 2 March 2025, 7:53 PM
Duration	9 mins 41 secs
Grade	**5.00** out of 5.00 (**100**%)

Question 1

Correct

Mark 1.00 out of 1.00

Suppose you have a dataset containing customer records for a company, where each record has a unique customer ID and an associated purchase amount. You want to sample customers for a survey, ensuring that both high-spending and low-spending customers are represented proportionally in the sample. Which sampling method would be the most appropriate in this scenario?

A. Reservoir sampling

B. Stratified sampling ✓

C. Random sampling

D. Convenience sampling

Your answer is correct.

Answer: B) Stratified sampling

Explanation:

In the given scenario, where we want to ensure that both high-spending and low-spending customers are represented proportionally in the sample, the most appropriate sampling method would be stratified sampling.

The correct answer is:
Stratified sampling

Question 2

Correct

Mark 1.00 out of 1.00

Which of the following techniques is used to increase the size and diversity of a training dataset by applying various transformations to the existing data?

A. Data Cleaning

B. Data Augmentation ✓

C. Feature Scaling

D. Data Labeling

Your answer is correct.

Answer:

B) Data Augmentation

The correct answer is:
Data Augmentation

Question 3

Correct

Mark 1.00 out of 1.00

Which of the following techniques is specifically designed to generate synthetic samples to address class imbalance?

- A. Undersampling
- B. Active learning
- C. SMOTE ✓
- D. Data Imputation

Your answer is correct.

Correct Answer: C) SMOTE

Explanation:

The Synthetic Minority Over-sampling Technique (SMOTE) is a resampling method that generates synthetic samples for the minority class rather than simply duplicating existing ones. It does this by interpolating between existing minority class samples to create new, similar instances. This helps improve class balance without causing overfitting due to repeated duplication. Other methods like undersampling (A) remove majority class samples, cost-sensitive learning (B) assigns different misclassification costs, and ensemble methods (D) combine models, but they do not generate synthetic data like SMOTE.

The correct answer is:
SMOTE

Question 4

Correct

Mark 1.00 out of 1.00

Which of the following techniques adjusts the range of numerical values in the dataset, often to a standard range, to improve model performance?

- A. Scaling ✓
- B. Discretization
- C. Imputation
- D. Deletion

Your answer is correct.

Answer:

A) Scaling

The correct answer is:
Scaling

Question 5

Correct

Mark 1.00 out of 1.00

In the context of handling missing data, what does the term "imputation" refer to?

A. Removing rows or columns with missing data

B. Replacing missing data with an estimated value derived from existing data. ✓

C. Both A & B

D. None of the above

Your answer is correct.

Answer: B) Filling missing data with a predicted or estimated value based on other data

The correct answer is:
Replacing missing data with an estimated value derived from existing data.