

# Module 09 – Quiz & Explanation

## 1. Which method is commonly used to detect outliers in numerical data?

- a) Min-Max Scaling
- b) Z-Score
- c) One-Hot Encoding
- d) Label Encoding

**Ans: b) Z-Score**

ব্যাখ্যা: Z-Score দিয়ে আমরা একটি value Mean থেকে কত দূরে আছে তা বুঝতে পারি। যদি একটি value 3 standard deviation-এর বেশি দূরে যায়, তাকে সাধারণত Outlier ধরা হয়। তাই outlier detection-এ Z-Score খুব কার্যকর।

## 2. What is the main purpose of outlier handling?

- a) Reduce dataset size
- b) Improve model accuracy
- c) Convert text to numbers
- d) Remove multicollinearity

**Ans: b) Improve model accuracy**

ব্যাখ্যা: Outlier মডেলকে ভুল দিকে নিয়ে যায় এবং prediction-এ noise আনে। Outlier remove/transform করলে model Stability ও Accuracy বাড়ে—এটাই মূল উদ্দেশ্য।

## 3. Polynomial Features help the model to capture:

- a) Linear relationships only
- b) Non-linear relationships
- c) Missing values
- d) Duplicates

**Ans: b) Non-linear relationships**

ব্যাখ্যা: Polynomial Features (যেমন  $x^2$ ,  $x^3$ ) ডেটার non-linear pattern ধরতে সাহায্য করে। যেমন house price dataset-এ size<sup>2</sup> relationship ধরতে polynomial terms কাজে লাগে।

#### **4. Which transformation is used to reduce skewness in data?**

- a) Log Transformation
- b) Standard Scaling
- c) One-Hot Encoding
- d) Binning

**Ans: a) Log Transformation**

ব্যাখ্যা: Skewed ডিস্ট্রিবিউশনকে normal-এর কাছে আনার জন্য Log Transformation খুব কার্যকর। Income বা sales এর মতো heavy-tailed ডাটায় এটি বেশি ব্যবহার হয়।

#### **5. Domain-Driven Feature Engineering means:**

- a) Using random transformations
- b) Using SME (subject matter expertise)
- c) Automatically generated features
- d) Unsupervised clustering

**Ans: b) Using SME (subject matter expertise)**

ব্যাখ্যা: Domain knowledge দিয়ে নতুন meaningful feature তৈরি করাই Domain-Driven Feature Engineering। যেমন finance-এ "Debt-to-Income Ratio" বানানো।

#### **6. In preprocessing pipelines, what is the correct order?**

- a) Train model → Scale data → Handle missing values
- b) Scale data → Train model → Handle outliers
- c) Handle missing values → Handle outliers → Scale data
- d) Train model → Encode categories → Scale data

**Ans: c) Handle missing values → Handle outliers → Scale data**

ব্যাখ্যা: প্রথমে missing value ঠিক করতে হয়, তারপর outlier handle, শেষে scaling। এভাবে pipeline পরিষ্কার থাকে এবং downstream model stable হয়।

## **7. What does Binning help with in feature engineering?**

- a) Converting numeric values into categories
- b) Increasing model complexity
- c) Creating outliers
- d) Reducing dimensionality

**Ans: a) Converting numeric values into categories**

ব্যাখ্যা: Continuous numeric data কে ranges বা buckets-এ ভাগ করাই Binning। যেমন age → (0–18, 19–35, 36–60)। এতে interpretability বাড়ে।

## **8. Why do we create interaction features (like $x_1 * x_2$ )?**

- a) To reduce model training time
- b) To capture combined effects of variables
- c) To replace missing values
- d) To detect multicollinearity

**Ans: b) To capture combined effects of variables**

ব্যাখ্যা: Interaction features দুইটা variable মিলে outcome-এ কী প্রভাব ফেলছে তা ধরতে পারে। যেমন “Age \* Income” হ্যাতো loan approval-এ প্রভাব ফেলে।

## **9. Capping outliers (Winsorization) means:**

- a) Removing outliers completely
- b) Replacing extreme values with boundary values
- c) Converting outliers to missing values
- d) Ignoring all outliers

**Ans: b) Replacing extreme values with boundary values**

ব্যাখ্যা: Winsorization পদ্ধতিতে extreme high/low values কে 95th বা 5th percentile boundary-তে নিয়ে আসা হয়। এতে distribution stable থাকে।

## **10. In a preprocessing pipeline, why is automation important?**

- a) It reduces dataset columns
- b) It prevents data leakage
- c) It adds random noise
- d) It skips missing values handling

**Ans: b) It prevents data leakage**

ব্যাখ্যা: Pipeline ব্যবহার করলে একই steps train ও test সেটে ঠিকমত apply হয়, ফলে data leakage কমে। Automation consistency নিশ্চিত করে।