Exercise: Chapter 1 – Capstone Project

✅ A. Multiple Choice Questions (MCQs)

1. What is the **first step** in an AI project cycle?
   - a) Feature Definition
   - b) Model Evaluation
   - c) Problem Definition
   - d) Deployment

   → **Correct: c)**
2. Which of the following is **not** one of the five main types of AI questions?
   - a) Which group?
   - b) What happened yesterday?
   - c) Is this unusual?
   - d) How much or how many?

   → **Correct: b)**
3. The Design Thinking process includes:
   - a) Predict, Analyze, Act
   - b) Empathize, Define, Ideate, Prototype, Test
   - c) Understand, Deploy, Evaluate
   - d) Create, Build, Validate

   → **Correct: b)**
4. What does the acronym **RMSE** stand for?
   - a) Root Matrix Scaled Error
   - b) Root Mean Squared Error
   - c) Real Model Square Evaluation
   - d) Rounded Mean Sampling Error

   → **Correct: b)**
5. Which method is more **accurate than train-test split** for small datasets?
   - a) Confusion Matrix
   - b) Reinforcement Learning
   - c) Cross-validation
   - d) Neural Nets

   → **Correct: c)**

✍️ B. Fill in the Blanks

1. The final project in AI curriculum that integrates all learning is known as the _____ **Project**.
   → Capstone
2. In regression problems, the output is a _____ **value**.
   → continuous

3. The process of breaking a large problem into smaller parts is called _____ **decomposition**.
   → problem
4. A model's accuracy in regression can be evaluated using _____, **MSE,** **and** **MAE**.
   → RMSE
5. The five stages of Design Thinking are **Empathize, Define, Ideate, Prototype,** **and** _____.
   → Test

Questions and Answers with Examples

❓ Q1. What is a Capstone Project in Artificial Intelligence? Give one example.

Answer:
A Capstone Project is a final project that integrates all the knowledge and skills a student has acquired in the AI curriculum. It involves solving a real-world problem using the AI Project Cycle.

Example:
Predicting student results based on study hours, attendance, and internal assessments using a regression model.

❓ Q2. What are the six steps in the AI Project Cycle? Explain with an example.

Answer:
The six steps in the AI Project Cycle are:

1. **Problem Definition** – Understanding the problem.
2. **Data Gathering** – Collecting relevant data.
3. **Feature Definition** – Selecting important variables.
4. **Model Construction** – Building a predictive or classification model.
5. **Evaluation & Refinement** – Testing and improving the model.
6. **Deployment** – Implementing the model in the real world.

Example:
If we want to **predict movie ticket prices**, we:

- Understand factors affecting price (problem),
- Collect historical data on movies,

- Define features like genre, duration, release month,
- Train a model using regression,
- Evaluate accuracy using RMSE,
- Deploy it for users to forecast future ticket costs.

❓ Q3. How can we identify whether a problem is suitable for AI?

Answer:
We can use AI only when the problem has a **recognizable pattern** in the data. If no pattern exists, AI cannot be applied effectively.

Example:
Predicting the **next word in a sentence** (pattern exists) is suitable for AI. Predicting **random lottery numbers** (no pattern) is not.

❓ Q4. What is Design Thinking, and how does it help in AI projects?

Answer:
Design Thinking is a method to solve complex problems through empathy and user-centered design. It has five stages: **Empathize, Define, Ideate, Prototype, Test**.

Example:
If you're building an AI-based fitness app:

- Empathize with users who want to lose weight,
- Define their needs (personalized plans),
- Ideate features (diet suggestions, activity tracker),
- Build a prototype with dummy data,
- Test it with feedback from real users.

❓ Q5. What is the difference between Regression and Classification? Give examples.

Answer:

- **Regression** predicts a **numerical value**.
- **Classification** predicts a **category or label**.

Examples:

- Regression: Predicting **house prices** based on size and location.
- Classification: Identifying whether an **email is spam or not**.

❓ Q6. What is Train-Test Split and how is it used in AI modeling?

Answer:
Train-Test Split is a technique to evaluate AI models. The dataset is split into a **training set** (to build the model) and a **test set** (to evaluate it).

Example in Python:

```
python
CopyEdit
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
```

❓ Q7. What is RMSE and how is it useful in evaluating models? Give an example.

Answer:
RMSE (Root Mean Squared Error) measures how much predicted values differ from actual values. Lower RMSE means better accuracy.

Example:
If a model predicts student marks:

- Actual marks: [90, 85, 70]
- Predicted: [88, 80, 75]
  RMSE will show how far off the predictions are, on average.

❓ Q8. What is Cross-Validation and why is it preferred over Train-Test split sometimes?

Answer:
Cross-validation divides data into multiple folds and rotates them between training and testing. It gives a **more reliable model evaluation**, especially for small datasets.

Example:
In 5-fold cross-validation, the model is trained and tested **5 times** using different parts of the data each time, ensuring no part is left untested.

❓ Q9. What kind of real-life problems can be solved using AI Capstone Projects? Give at least two examples.

Answer:
AI Capstone Projects solve real-life issues by predicting or classifying data.

Examples:

1. **Healthcare** – Predicting whether a patient will develop diabetes based on health indicators.
2. **Agriculture** – Using AI to detect plant diseases from leaf images.

Long Answer Questions (With Examples)

❓ Q1. Explain the AI Project Cycle in detail with a real-life example.

Answer:
The **AI Project Cycle** is a structured way to solve real-world problems using Artificial Intelligence. It consists of six main steps:

1. **Problem Definition:**
   Understand the problem and determine whether it has a pattern. If there's no pattern, AI cannot help.
   Example: Predicting crop yield based on rainfall, soil type, and temperature.
2. **Data Gathering:**
   Collect data relevant to the problem. This may include past records, surveys, sensors, etc.
   Example: Gather past 10 years' data on crop yield, weather, and farming practices.
3. **Feature Definition:**
   Identify important variables (features) that influence the output.
   Example: For crop yield prediction, features could be rainfall (mm), fertilizer quantity, and seed quality.
4. **Model Construction:**
   Use machine learning algorithms to build a model that can learn patterns from the data.
   Example: Train a regression model to predict yield using historical data.
5. **Evaluation & Refinement:**
   Test the model's performance using metrics like RMSE or accuracy, and improve it by changing features or model types.
   Example: If RMSE is high, refine the model by adding new features like pest control usage.
6. **Deployment:**
   Use the trained model in real-world applications.

Example: Deploy the model in a mobile app for farmers to predict yield based on inputs.

❓ Q2. What is Design Thinking? Describe its stages and explain how it can be used to solve an AI problem.

Answer:
Design Thinking is a human-centered approach used for solving complex and undefined problems creatively and efficiently. It has five stages:

1. **Empathize:**
   Understand the users and their problems by observing and interacting.
   Example: Interview farmers to learn the challenges they face in pest control.
2. **Define:**
   Clearly state the problem using insights gathered.
   Example: "Farmers cannot detect early signs of pest attacks on crops."
3. **Ideate:**
   Brainstorm possible solutions with creativity and without judgment.
   Example: Ideas could include AI-based image recognition, mobile notifications, drone analysis, etc.
4. **Prototype:**
   Build a simple version of the selected idea to see how it works.
   Example: Develop a basic app that uses photos of leaves to identify pest issues.
5. **Test:**
   Try the prototype with real users, collect feedback, and improve the solution.
   Example: Give the app to a few farmers and improve it based on their input.

This method ensures that AI solutions are practical, user-friendly, and targeted.

❓ Q3. Describe Train-Test Split and Cross-Validation techniques in AI. When and why should each be used? Provide code examples.

Answer:
Both Train-Test Split and Cross-Validation are methods to evaluate AI model performance.

1. Train-Test Split:
Divides the dataset into two parts:

- **Training Set:** To train the model.
- **Test Set:** To evaluate the model's predictions.

When to use:

- When you have a large dataset.
- Fast evaluation required.

Example in Python:

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
```

2. Cross-Validation: Splits data into **k subsets (folds)** and runs k experiments where each fold is used once as a test set.

When to use:

- When you have a small dataset.
- You need a more **reliable** and **less biased** performance score.

Example in Python:

```
from sklearn.model_selection import cross_val_score
from sklearn.ensemble import RandomForestRegressor
scores = cross_val_score(RandomForestRegressor(), X, y, scoring='neg_mean_absolute_error', cv=5)
print("MAE:", -1 * scores.mean())
```

Conclusion:

- **Train-Test Split** is faster, suitable for big data.
- **Cross-Validation** is more accurate and preferred when data is limited.

❓ Q4. What are RMSE and MSE? How are they calculated and interpreted? Give examples.

Answer:
Both RMSE (Root Mean Squared Error) and MSE (Mean Squared Error) are metrics used to evaluate the accuracy of regression models.

---

1. MSE (Mean Squared Error):
It is the average of the squared differences between actual and predicted values.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

2. RMSE (Root Mean Squared Error):
It is the square root of MSE and provides the error in the same units as the target variable.

$$RMSE = \sqrt{MSE}$$

Example:

- Actual: [2, 4, 6]
- Predicted: [2.2, 3.8, 5.9]

$$MSE = \frac{(2-2.2)^2 + (4-3.8)^2 + (6-5.9)^2}{3} = \frac{0.04 + 0.04 + 0.01}{3} = 0.03$$

$$RMSE = \sqrt{0.03} \approx 0.173$$

Interpretation:

- Lower RMSE/MSE = better model.
- If RMSE is high, the model predictions are far from actual values.