Viva Questions & Answers - ACL Experiments

1. Build a Simple ML Model on Cloud using Iris Dataset

Q: What is the Iris dataset?

A: A classic dataset with 150 samples of iris flowers used to classify species based on sepal/petal length and width.

Q: Which cloud platform did you use and why?

A: Example: AWS SageMaker, because it offers easy model building, training, and deployment.

Q: What algorithm did you use for classification?

A: Logistic Regression, Decision Tree, or any basic classifier.

Q: How did you evaluate the model?

A: Using accuracy, confusion matrix, or cross-validation.

2. Implement Data Preprocessing using AWS SageMaker

Q: Why is preprocessing important?

A: It cleans and prepares data, improving model performance.

Q: What techniques did you apply?

A: Handling missing values, normalization, encoding categorical variables.

Q: How is preprocessing done in SageMaker?

A: Using Jupyter notebooks with Pandas/Scikit-learn or built-in pipelines.

3. Implement Data Visualization using AWS SageMaker

Q: Why do we visualize data?

A: To understand data distribution, detect outliers, and reveal patterns.

Q: Which libraries did you use?

A: Matplotlib, Seaborn, or built-in SageMaker Studio tools.

Q: What types of charts did you use?

A: Histograms, scatter plots, box plots, or heatmaps.

4. Implement Image Binary Classification using Cloud

Q: What is binary classification?

A: Classification with two classes (e.g., cat vs. dog, healthy vs. unhealthy leaf).

Q: What cloud platform and tool did you use?

A: AWS SageMaker or Google Vertex AI using CNN.

Q: How did you preprocess the images?

A: Resizing, normalization, and data augmentation.

5. Natural Language Preprocessing with AWS

Q: Which AWS service is used for NLP?

A: Amazon Comprehend.

Q: What tasks can Comprehend perform?

A: Sentiment analysis, entity recognition, key phrase extraction.

Q: What input format does Comprehend require?

A: Plain text (UTF-8), often in JSON or string format.

6. Creating a Linear Regression Model using Amazon/Colab

Q: What is linear regression?

A: A statistical method to model the relationship between variables ($y = mx + c$).

Q: Where did you build your model?

A: Amazon Q Developer or Google Colab.

Q: What evaluation metric did you use?

A: Mean Squared Error (MSE) or R^2 score.

7. Implement Real-Time Data Streaming (e.g., Weather/Stock/Traffic)

Q: What is real-time data processing?

A: Processing data instantly as it's generated.

Q: Which service did you use for streaming?

A: AWS Kinesis or Google DataFlow.

Q: How did you use the data for predictions?

A: Passed it to a live SageMaker model or stored in real-time dashboards.

8. Build Serverless AI API Using AWS Lambda & API Gateway

Q: What is AWS Lambda?

A: A serverless service that runs code in response to events.

Q: What is API Gateway?

A: A tool to create RESTful APIs that trigger Lambda functions.

Q: How is serverless beneficial?

A: No need to manage servers, auto-scales, and cost-effective.