WB Board Class 12 Artificial Intelligence Question Paper with Solutions(Memory Based)

Time Allowed :3 Hour	**Maximum Marks :**60	**Total Questions :**24

General Instructions

Read the following instructions very carefully and strictly follow them:

- Answers to this Paper must be written on the paper provided separately.
- You will not be allowed to write during the first 15 minutes
- This time is to be spent in reading the question paper.
- The time given at the head of this Paper is the time allowed for writing the answers,
- The paper has four Sections.
- Section A is compulsory - All questions in Section A must be answered.
- You must attempt one question from each of the Sections B, C and D and one other question from any Section of your choice.

1. Explain the 10 stages of the Foundational Methodology for Data Science.

Correct Answer: The Foundational Methodology for Data Science consists of 10 stages: Business Understanding, Analytic Approach, Data Requirements, Data Collection, Data Understanding, Data Preparation, Modeling, Evaluation, Deployment, and Feedback.

Solution: Concept: The Foundational Methodology for Data Science is a structured framework that guides data scientists through the lifecycle of a data science project — from problem definition to deployment and continuous improvement.
Step 1: Business Understanding
This stage defines the problem from a business perspective:

- Identify objectives and goals
- Understand stakeholders' needs
- Define success criteria

Step 2: Analytic Approach
Determine the appropriate analytical technique:

- Classification, regression, clustering, etc.
- Choose methods based on problem type

Step 3: Data Requirements
Specify the type of data needed:

- Structured or unstructured data
- Data sources and formats

Step 4: Data Collection
Gather the required data from:

- Databases, APIs, surveys, logs
- Internal and external sources

Step 5: Data Understanding
Explore and analyze the collected data:

- Identify patterns and anomalies
- Perform exploratory data analysis (EDA)

Step 6: Data Preparation
Clean and transform data for modeling:

- Handle missing values
- Normalize and encode variables
- Feature engineering

Step 7: Modeling
Build predictive or analytical models:

- Select algorithms
- Train models using prepared data

Step 8: Evaluation
Assess model performance:

- Use validation metrics (accuracy, precision, RMSE)
- Compare multiple models

Step 9: Deployment
Implement the model in real-world systems:

- Integrate into applications or dashboards
- Enable real-time or batch predictions

Step 10: Feedback and Monitoring
Continuously improve the solution:

- Monitor model performance
- Collect user feedback
- Retrain models as needed

Quick Tip

A successful data science project moves from **problem understanding** to **model deployment** and ends with **continuous feedback and improvement**.

---

2. Why is the "Business Understanding" stage considered the most critical phase of an AI project?

Correct Answer: The Business Understanding stage is most critical because it defines the problem, aligns AI goals with business objectives, and determines the success criteria, ensuring the project delivers meaningful and practical value.

Solution: Concept: The Business Understanding stage lays the foundation of an AI project. It ensures that the technical work aligns with real-world needs, preventing wasted effort on irrelevant or poorly defined problems.
Step 1: Defines the Real Problem
This stage clarifies:

- What problem needs to be solved
- Why it matters to the organization

Without proper understanding, AI solutions may solve the wrong problem.
Step 2: Aligns AI with Business Goals
It ensures:

- AI outcomes support strategic objectives
- Measurable business value (profit, efficiency, user experience)

This prevents purely academic or impractical models.
Step 3: Defines Success Metrics
Clear KPIs are established, such as:

- Revenue growth
- Cost reduction
- Accuracy or customer satisfaction

These metrics guide evaluation later in the project.
Step 4: Guides the Entire AI Lifecycle
Decisions in later stages depend on this phase:

- Choice of data and models
- Evaluation criteria
- Deployment strategy

A weak foundation leads to flawed downstream decisions.
Step 5: Reduces Risk and Resource Waste
Proper business understanding helps:

- Avoid unnecessary data collection
- Prevent costly development mistakes
- Ensure stakeholder alignment

Quick Tip

If the problem is misunderstood, even the best AI model will fail.
Strong business understanding ensures AI delivers real value.

---

3. Define the 5 Vs of Big Data: Volume, Velocity, Variety, Veracity, and Value.

Correct Answer: The 5 Vs of Big Data describe its key characteristics: Volume (amount of data), Velocity (speed of data generation), Variety (different data types), Veracity (data reliability), and Value (usefulness of data insights).

Solution: Concept: Big Data refers to extremely large and complex datasets that cannot be processed using traditional data processing tools. The 5 Vs framework helps describe the core characteristics that define Big Data systems and challenges.
Step 1: Volume
Volume refers to the massive amount of data generated and stored:

- Measured in terabytes, petabytes, or exabytes
- Generated from sources like social media, sensors, and transactions

Step 2: Velocity
Velocity describes the speed at which data is generated and processed:

- Real-time or near real-time data streams
- Examples: stock markets, IoT devices, online activity

Step 3: Variety
Variety refers to different forms of data:

- Structured (databases, tables)
- Semi-structured (JSON, XML)
- Unstructured (images, videos, text)

Step 4: Veracity
Veracity represents the quality and reliability of data:

- Noise, inconsistencies, or missing values
- Importance of data cleaning and validation

Step 5: Value
Value refers to the meaningful insights derived from data:

- Turning raw data into actionable intelligence
- Supporting better decision-making and innovation

Quick Tip

Big Data is not just about size — it's about handling large, fast, diverse, reliable data and extracting real value from it.

---

4. Describe the five stages of Design Thinking: Empathize, Define, Ideate, Prototype, and Test.

Correct Answer: The five stages of Design Thinking are Empathize (understand users), Define (clarify the problem), Ideate (generate solutions), Prototype (build models), and Test (evaluate and refine solutions).

Solution: Concept: Design Thinking is a human-centered problem-solving approach that focuses on understanding users, challenging assumptions, and creating innovative solutions through iterative development.
Step 1: Empathize
This stage focuses on understanding users deeply:

- Conduct interviews and observations
- Understand user needs, emotions, and challenges

The goal is to gain real user insights.
Step 2: Define
In this stage, the problem is clearly articulated:

- Analyze insights gathered during empathy
- Formulate a clear problem statement

This ensures the team solves the right problem.
Step 3: Ideate
The ideation phase involves generating creative solutions:

- Brainstorming sessions
- Encouraging innovative and diverse ideas

Quantity and creativity of ideas are emphasized.
Step 4: Prototype
Here, ideas are turned into tangible forms:

- Create low-cost models or mockups
- Build wireframes, sketches, or demos

Prototypes make abstract ideas testable.
Step 5: Test
The final stage evaluates the prototype:

- Gather user feedback
- Identify improvements or redesign needs

This stage is iterative and may lead back to earlier stages.

Quick Tip

Design Thinking is iterative — learn from users, build fast prototypes, test often, and refine continuously.

---

5. Differentiate between Supervised and Unsupervised Learning with real-world examples.

Correct Answer: Supervised learning uses labeled data to train models for prediction tasks, while unsupervised learning works with unlabeled data to discover patterns or groupings without predefined outputs.

Solution: Concept: Machine learning algorithms are broadly categorized into supervised and unsupervised learning based on the availability of labeled data and the learning objective.
Step 1: Supervised Learning
Supervised learning involves training a model using labeled datasets:

- Each input has a known output (label)
- The model learns to map inputs to outputs

Examples:

- Email spam detection (spam vs non-spam)
- House price prediction based on features
- Medical diagnosis using patient data

Step 2: Unsupervised Learning
Unsupervised learning works with unlabeled data:

- No predefined outputs
- The model identifies hidden patterns or structures

Examples:

- Customer segmentation in marketing
- Grouping similar news articles
- Anomaly detection in fraud analysis

Step 3: Key Differences

- **Data:** Supervised uses labeled data; unsupervised uses unlabeled data

- **Goal:** Prediction vs pattern discovery
- **Output:** Known outcomes vs hidden structures

Step 4: Use Case Perspective

- Use supervised learning when historical labeled data exists
- Use unsupervised learning for exploration and insights

Quick Tip

Supervised = Learn with answers.
Unsupervised = Discover patterns without answers.

---

6. Explain the concept of Linear Regression and the role of the Mean Squared Error (MSE) loss function.

Correct Answer: Linear Regression is a supervised learning method that models the relationship between variables using a straight line. The Mean Squared Error (MSE) loss function measures the average squared difference between predicted and actual values, guiding the model to minimize prediction errors.

Solution: Concept: Linear Regression is one of the simplest and most widely used machine learning algorithms for predicting a continuous output based on input features. It assumes a linear relationship between independent variables and the dependent variable.
Step 1: What is Linear Regression?
Linear regression models the relationship using a linear equation:

$$y = mx + c$$

where:

- y = predicted output
- x = input feature
- m = slope (weight)
- c = intercept (bias)

For multiple features, the equation extends to multiple linear regression.
Step 2: Goal of Linear Regression
The objective is to:

- Find the best-fitting line
- Minimize the difference between predicted and actual values

This difference is called the **error** or **residual**.
Step 3: Mean Squared Error (MSE) Loss Function
MSE is used to quantify prediction error:

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

where:

- y_i = actual value
- $\hat{y}_i$ = predicted value
- n = number of samples

Step 4: Role of MSE in Training
MSE plays a critical role by:

- Penalizing larger errors more heavily (due to squaring)
- Providing a smooth, differentiable loss function
- Enabling optimization using gradient descent

Step 5: Why MSE is Preferred

- Simple and mathematically convenient
- Works well with linear models
- Helps achieve a unique optimal solution

Quick Tip

Linear Regression finds the best-fit line, and MSE measures how far predictions are from actual values — lower MSE means better model performance.

---

7. What is the Gradient Descent algorithm and how does it help in model optimization?

Correct Answer: Gradient Descent is an optimization algorithm that iteratively adjusts model parameters to minimize a loss function by moving in the direction of the negative gradient, helping models achieve optimal performance.

Solution: Concept: Gradient Descent is a widely used optimization algorithm in machine learning and deep learning. It helps find the optimal values of model parameters (weights and biases) by minimizing a loss function, thereby improving prediction accuracy.
Step 1: What is Gradient Descent?
Gradient Descent is an iterative optimization method that:

- Calculates the gradient (slope) of the loss function
- Updates parameters in the opposite direction of the gradient

This ensures movement toward the minimum loss.
Step 2: Basic Update Rule
The parameter update is given by:

$$\theta = \theta - \alpha \cdot \nabla J(\theta)$$

where:

- θ = model parameters
- α = learning rate
- $\nabla J(\theta)$ = gradient of the loss function

Step 3: Role in Model Optimization
Gradient Descent helps by:

- Reducing prediction errors
- Finding parameter values that minimize loss
- Improving model accuracy over iterations

Step 4: Learning Rate Importance
The learning rate controls step size:

- Too large → overshooting the minimum
- Too small → slow convergence

Choosing the right value is critical for optimization.
Step 5: Variants of Gradient Descent

- **Batch Gradient Descent** — uses entire dataset
- **Stochastic Gradient Descent (SGD)** — updates per sample
- **Mini-batch Gradient Descent** — balance of speed and stability

Quick Tip

Gradient Descent works like walking downhill to reach the lowest point — each step reduces the model's error.

---

8. Explain the purpose of the Train-Test Split technique in evaluating model performance.

Correct Answer: The Train-Test Split divides data into separate training and testing sets to evaluate how well a model generalizes to unseen data, helping detect overfitting and ensuring reliable performance assessment.

Solution: Concept: The Train-Test Split is a fundamental evaluation technique in machine learning used to assess a model's ability to perform on new, unseen data. It prevents overly optimistic results by separating training and evaluation data.
Step 1: What is Train-Test Split?
The dataset is divided into two parts:

- **Training set** — used to train the model
- **Testing set** — used to evaluate performance

A common split is 80:20 or 70:30.
Step 2: Purpose of Training Data
The training data helps:

- Learn patterns and relationships
- Adjust model parameters

The model sees this data during learning.
Step 3: Purpose of Testing Data
The testing data is:

- Completely unseen during training
- Used for unbiased evaluation

It simulates real-world performance.
Step 4: Detecting Overfitting
Train-test split helps identify:

- High training accuracy but low test accuracy → overfitting
- Similar performance → good generalization

Step 5: Ensuring Model Reliability
This technique:

- Provides realistic performance estimates
- Prevents data leakage
- Builds trustworthy models

Quick Tip

Always test models on unseen data — a model that performs well only on training data is not reliable.

---

9. Explain the biological motivation behind Artificial Neural Networks (ANN).

Correct Answer: Artificial Neural Networks are inspired by the structure and functioning of the human brain, where interconnected neurons process information through signals and learning occurs by adjusting connection strengths.

Solution: Concept: Artificial Neural Networks (ANNs) are computational models inspired by the biological neural networks of the human brain. They mimic how neurons communicate and learn from experience by adjusting connections based on inputs and outputs.
Step 1: Biological Neurons in the Brain
In the human brain:

- Neurons are the basic processing units
- They receive signals through dendrites
- Signals are transmitted via axons to other neurons

Learning occurs through changes in synaptic strengths.
Step 2: Artificial Neuron Analogy
An artificial neuron mimics this structure:

- Inputs represent incoming signals
- Weights represent synaptic strengths
- Activation function determines output

The output is passed to the next layer of neurons.
Step 3: Learning Mechanism
In biological systems:

- Learning occurs by strengthening or weakening synapses

Similarly, in ANNs:

- Weights are updated during training
- Algorithms like backpropagation adjust connections

Step 4: Network Structure
Both biological and artificial networks:

- Contain interconnected nodes
- Process information in parallel
- Learn from experience

ANNs use layers such as input, hidden, and output layers.
Step 5: Key Motivation
The main inspiration is:

- Replicating human-like learning and pattern recognition
- Solving complex problems like vision, speech, and language understanding

Quick Tip

ANNs are brain-inspired models — artificial neurons simulate how biological neurons learn through interconnected signals and adaptive strengths.

10. Define Computer Vision and list its primary pipeline stages (Acquisition, Preprocessing, Feature Extraction).

Correct Answer: Computer Vision is a field of AI that enables machines to interpret and analyze visual data. Its primary pipeline stages include Image Acquisition, Preprocessing, and Feature Extraction.

Solution: Concept: Computer Vision is a branch of Artificial Intelligence that allows computers to understand and interpret images and videos, similar to human vision. It enables applications such as face recognition, object detection, and medical imaging analysis.
Step 1: Image Acquisition
This is the first stage where visual data is captured:

- Using cameras, sensors, or imaging devices
- Sources include photos, video streams, or scanned documents

The goal is to obtain raw visual input for processing.
Step 2: Preprocessing
In this stage, the acquired image is prepared for analysis:

- Noise reduction and filtering
- Image resizing or normalization
- Contrast enhancement or grayscale conversion

Preprocessing improves image quality and consistency.
Step 3: Feature Extraction
This stage identifies meaningful patterns from images:

- Detecting edges, textures, or shapes
- Extracting key points or descriptors

These features are later used for tasks like classification or detection.

Quick Tip

Computer Vision pipeline: Capture the image → Clean and enhance it → Extract meaningful features for AI models.

---

11. Explain the Turing Test and its significance in measuring machine intelligence.

Correct Answer: The Turing Test, proposed by Alan Turing, evaluates a machine's intelligence based on its ability to exhibit human-like conversational behavior indistinguishable from a human. It is significant as an early benchmark for assessing artificial intelligence.

Solution: Concept: The Turing Test is a foundational concept in artificial intelligence introduced by Alan Turing in 1950. It was designed to determine whether a machine can demonstrate intelligent behavior comparable to that of a human.

Step 1: What is the Turing Test?
The Turing Test involves three participants:

- A human judge
- A human participant
- A machine

The judge interacts with both through text-based communication without knowing which is which.
Step 2: Test Procedure

- The judge asks questions to both entities
- If the judge cannot reliably distinguish the machine from the human

The machine is said to have passed the Turing Test.
Step 3: Significance in AI
The Turing Test is important because:

- It provides an early definition of machine intelligence
- Focuses on behavior rather than internal mechanisms
- Encourages development of human-like AI interaction

Step 4: Limitations
Despite its importance:

- It measures imitation, not true understanding
- A machine may appear intelligent without real reasoning

Step 5: Modern Relevance
Today, the Turing Test:

- Serves as a historical benchmark
- Inspires research in conversational AI and chatbots

Quick Tip

If a machine can converse indistinguishably from a human, it passes the Turing Test — a classic milestone in AI history.

---

12. Discuss the major ethical issues in AI, such as Bias, Privacy, and Accountability.

Correct Answer: Major ethical issues in AI include Bias (unfair or discriminatory outcomes), Privacy (misuse of personal data), and Accountability (lack of clarity on responsibility for AI decisions).

Solution: Concept: As Artificial Intelligence becomes widely integrated into society, it raises significant ethical concerns. Ensuring fairness, protecting personal data, and defining responsibility are crucial challenges in responsible AI development.
Step 1: Bias in AI
Bias occurs when AI systems produce unfair or discriminatory outcomes due to:

- Biased training data
- Skewed sampling or historical inequalities

Examples include biased hiring tools or facial recognition errors across demographics.
Step 2: Privacy Concerns
AI systems often rely on large amounts of personal data:

- Risk of data misuse or unauthorized access
- Surveillance concerns
- Data collection without informed consent

This raises serious issues regarding user rights and data protection.
Step 3: Accountability
Accountability refers to responsibility for AI decisions:

- Who is liable if an AI system fails?
- Developers, organizations, or users?

Lack of clear accountability complicates legal and ethical frameworks.
Step 4: Broader Ethical Implications
These issues can lead to:

- Loss of public trust in AI
- Legal and regulatory challenges
- Social inequality if not addressed

Step 5: Importance of Responsible AI
Addressing ethical concerns requires:

- Fair and transparent algorithms
- Strong data protection policies
- Clear governance and regulations

Quick Tip

Responsible AI must be fair (no bias), respectful (protect privacy), and answerable (clear accountability).

---

13. What is a Confusion Matrix, and how do you calculate Accuracy, Precision, and Recall from it?

Correct Answer: A Confusion Matrix is a table used to evaluate classification models by comparing predicted and actual values. Accuracy, Precision, and Recall are calculated using True Positives, True Negatives, False Positives, and False Negatives derived from it.

Solution: Concept: A Confusion Matrix is a performance evaluation tool for classification models. It summarizes prediction results by comparing actual labels with predicted labels, helping analyze model strengths and weaknesses.

Step 1: Structure of a Confusion Matrix

A binary classification confusion matrix contains four components:

- **True Positive (TP):** Correctly predicted positive cases
- **True Negative (TN):** Correctly predicted negative cases
- **False Positive (FP):** Incorrectly predicted positive (Type I error)
- **False Negative (FN):** Incorrectly predicted negative (Type II error)

Step 2: Accuracy

Accuracy measures overall correctness of the model:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

It shows the proportion of total correct predictions.

Step 3: Precision

Precision measures prediction reliability for positive class:

$$\text{Precision} = \frac{TP}{TP + FP}$$

It answers: *Out of predicted positives, how many were correct?*

Step 4: Recall (Sensitivity)

Recall measures the model's ability to detect actual positives:

$$\text{Recall} = \frac{TP}{TP + FN}$$

It answers: *Out of actual positives, how many were correctly identified?*

Step 5: Why These Metrics Matter

- Accuracy works well with balanced datasets
- Precision is important when false positives are costly
- Recall is critical when missing positives is dangerous (e.g., medical diagnosis)

Quick Tip

From the confusion matrix:
Accuracy = overall correctness,
Precision = correctness of positive predictions,
Recall = ability to detect actual positives.