

**Course Name – Machine Learning**

**Course Code - BCA57204 (T)**

**Multiple Choice Type Questions**

1.

**(i) Which of the following best defines Machine Learning?**

- |   |  |
|---|--|
| a) Hard-coding decisions into software      | b) Making computers intelligent without data |
| c) Programming computers to learn from data | d) Installing AI tools in a system           |

**(ii) Machine Learning is a subfield of:**

- |                    |                            |
|--------------------|----------------------------|
| a) Data Structures | b) Artificial Intelligence |
| c) Cyber Security  | d) Software Engineering    |

**(iii) Which of the following is NOT a type of Machine Learning?**

- |                  |                 |
|------------------|-----------------|
| a) Supervised    | b) Unsupervised |
| c) Reinforcement | d) Directed     |

**(iv) What is the primary goal of predictive analytics?**

- |                                 |                               |
|---------------------------------|-------------------------------|
| a) To understand past behaviors | b) To predict future outcomes |
| c) To store big data            | d) To clean datasets          |

**(v) Which technique is typically used in descriptive analytics?**

- |                       |                   |
|-----------------------|-------------------|
| a) Regression         | b) Classification |
| c) Data summarization | d) Reinforcement  |

**(vi) Which of the following best describes Big Data?**

- |  |  |
|--|--|
| a) Small and structured datasets           | b) Data with large volume, variety, and velocity |
| c) Data that is only in spreadsheet format | d) Clean and noise-free data                     |

**(vii) Which Python library is commonly used for data visualization?**

- |               |                 |
|---------------|-----------------|
| a) NumPy      | b) Matplotlib   |
| c) TensorFlow | d) Scikit-learn |

**(viii) Which of the following is NOT a Machine Learning application?**

- |                             |                       |
|-----------------------------|-----------------------|
| a) Email filtering          | b) Weather prediction |
| c) Solving algebra manually | d) Fraud detection    |

**(ix) In Machine Learning, training data is used to:**

- |                            |                         |
|----------------------------|-------------------------|
| a) Predict unknown results | b) Install new hardware |
| c) Build models            | d) Store logs           |

**(x) Which of the following is an example of supervised learning?**

- a) Clustering
- b) Linear Regression
- c) Market Basket Analysis
- d) Dimensionality Reduction

**(xi) Artificial Intelligence includes which of the following subfields?**

- a) Machine Learning
- b) Robotics
- c) Natural Language Processing
- d) All of the these

**(xii) What does the 'training' phase in Machine Learning do?**

- a) Runs system tests
- b) Builds a model from data
- c) Deletes old records
- d) Visualizes patterns

**(xiii) Which term refers to the ability of a machine to improve performance over time with more data?**

- a) Optimization
- b) Scalability
- c) Learning
- d) Memory

**(xiv) Which term best represents the comparison of ML with AI?**

- a) ML is a synonym for AI
- b) ML is independent of AI
- c) ML is a subset of AI
- d) ML is superior to AI

**(xv) What is the core idea of Machine Learning?**

- a) Writing specific rules for each task
- b) Storing large data
- c) Learning patterns from data
- d) Accessing internet data

**(xvi) Which of the following helps in identifying unknown patterns in data?**

- a) Supervised Learning
- b) Unsupervised Learning
- c) Reinforcement Learning
- d) Rule-based systems

**(xvii) Which metric is most commonly used to measure accuracy in ML?**

- a) Frequency
- b) Variance
- c) Accuracy Score
- d) File Size

**(xviii) Which one is NOT a Python library for data handling or visualization?**

- a) Pandas
- b) Seaborn
- c) NumPy
- d) CorelDraw

**(xix) Descriptive analytics is mostly concerned with:**

- a) What happened
- b) What will happen
- c) What should happen
- d) None of the these

**(xx) Which of the following is a challenge in Machine Learning?**

- a) Small data
- b) Manual computations
- c) Overfitting
- d) Static results

**(xxi) Which concept involves representing knowledge in ML systems?**

- a) Memory Mapping
- b) Knowledge Representation
- c) Bit Encoding
- d) Dataset Flattening

**(xxii) Which of the following is NOT a branch of Machine Learning?**

- a) Deep Learning
- b) Supervised Learning
- c) Unsupervised Learning
- d) Encrypted Learning

**(xxiii) What is the purpose of reinforcement learning?**

- a) To find data clusters
- b) To reward learning through actions
- c) To reduce file size
- d) To clean data

**(xxiv) Which tool is commonly used for predictive analytics?**

- a) MS Paint
- b) Excel Charts
- c) Regression Models
- d) Antivirus Software

**(xxv) Which aspect differentiates ML from traditional programming?**

- a) Uses only algorithms
- b) Learns from data
- c) Requires manual coding for every step
- d) Works offline only

**(xxvi) What is the main difference between supervised and unsupervised learning?**

- a) Supervised uses labeled data, Unsupervised doesn't
- b) Both use labels
- c) Unsupervised is faster
- d) Supervised doesn't need data

**(xxvii) Which learning type uses rewards and punishments?**

- a) Supervised
- b) Reinforcement
- c) Clustering
- d) Unsupervised

**(xxviii) What does a neural network mimic?**

- a) Computer hardware
- b) Human brain
- c) Animal cells
- d) Search engine

**(xxix) In which scenario is classification most appropriate?**

- a) Grouping customers
- b) Predicting temperature
- c) Identifying spam emails
- d) Calculating averages

**(xxx) Which is a key feature of supervised learning?**

- a) No training required
- b) Uses labeled data
- c) Unstructured output
- d) Uses only images

**(xxxi) Clustering is an example of:**

- a) Supervised
- b) Reinforcement
- c) Unsupervised
- d) Regression

**(xxxii) A decision tree is typically used for:**

- a) Regression only
- b) Supervised learning
- c) Clustering
- d) Data storage

**(xxxiii) Which of these is NOT a benefit of using neural networks?**

- a) Pattern recognition
- b) High transparency
- c) Adaptability
- d) Non-linearity

**(xxxiv) A key challenge in reinforcement learning is:**

- a) Overfitting
- b) Designing reward signals
- c) Noisy data
- d) Data labeling

**(xxxv) What is data preprocessing mainly used for?**

- a) Beautifying data
- b) Preparing clean, usable data
- c) Encrypting datasets
- d) Creating databases

**(xxxvi) What type of data is required for unsupervised learning?**

- a) Labeled
- b) Unlabeled
- c) Classified
- d) Binary

**(xxxvii) Which of the following best describes reinforcement learning?**

- a) Uses labeled data
- b) Requires real-time supervision
- c) Learns from interaction with environment
- d) Works offline only

**(xxxviii) Which algorithm is typically used for classification?**

- a) K-Means
- b) Linear Regression
- c) Decision Tree
- d) DBSCAN

**(xxxix) Which is most likely used for market segmentation?**

- a) Regression
- b) Reinforcement
- c) Clustering
- d) Classification

**(xl) How does supervised learning learn from data?**

- a) By creating labels
- b) By guessing
- c) From labeled examples
- d) With human interaction

**(xli) Which neural network layer adjusts weights during training?**

- a) Input
- b) Hidden
- c) Output
- d) Activation

**(xlii) Which is not a common ML application?**

- a) Language translation
- b) Image recognition
- c) Music composition
- d) Weather prediction

**(xliii) Reinforcement learning is inspired by:**

- a) Evolution
- b) Trial-and-error
- c) Predefined rules
- d) Probabilities

**(xliv) Which library is commonly used for ML in Python?**

- a) NumPy
- b) Pandas
- c) Scikit-learn
- d) OpenCV

**(xlv) Clustering tries to:**

- a) Classify
- b) Maximize variance
- c) Group similar data
- d) Predict labels

**(xlvi) The purpose of an activation function in a neural network is to:**

- a) Normalize data
- b) Add non-linearity
- c) Train weights
- d) Output labels

**(xlvii) In reinforcement learning, what represents a decision?**

- a) Action
- b) Reward
- c) State
- d) Agent

**(xlviii) A regression model predicts:**

- a) Labels
- b) Categories
- c) Continuous values
- d) Rewards

**(xlix) Which is an output of clustering?**

- a) Labels
- b) Classes
- c) Groups
- d) Trees

**(l) Business decisions can benefit from ML by:**

- a) Avoiding all risks
- b) Making assumptions
- c) Providing data-driven insights **(Correct Answer)**
- d) Replacing managers

**(li) What is the first step in training a machine learning model?**

- a) Model evaluation
- b) Feature engineering
- c) Data preparation
- d) Hyperparameter tuning

**(lii) Which algorithm is commonly used for regression problems?**

- a) Logistic Regression
- b) Linear Regression
- c) Decision Tree Classifier
- d) K-Means

**(liii) What is the purpose of the loss function during training?**

- a) To store data
- b) To reduce model complexity
- c) To calculate prediction error
- d) To increase accuracy

**(liv) Which data split is typically used to evaluate model generalization?**

- a) Training set
- b) Validation set
- c) Test set
- d) Development set

**(lv) Which of the following is a classification algorithm?**

- a) Linear Regression
- b) K-Means
- c) Logistic Regression
- d) PCA

**(lvi) What is the primary goal of training a machine learning model?**

- a) To make predictions
- b) To analyze data
- c) To clean data
- d) To visualize data

**(lvii) Which component adjusts weights in a neural network?**

- a) Backpropagation
- b) Activation function
- c) Epoch
- d) Loss function

**(lviii) What is an epoch in machine learning training?**

- a) A dataset
- b) One complete pass of the training dataset
- c) Loss function
- d) A testing method

**(lix) Which technique is used to evaluate model performance?**

- a) Overfitting
- b) Cross-validation
- c) Normalization
- d) One-hot encoding

**(lx) Which algorithm is used for clustering?**

- a) SVM
- b) Linear Regression
- c) K-Means
- d) Decision Tree

**(lxi) What does a high bias in a model usually indicate?**

- a) Overfitting
- b) Good generalization
- c) Underfitting
- d) Perfect model

**(lxii) Which step is essential before training a model?**

- a) Data visualization
- b) Data collection
- c) Data preprocessing
- d) Prediction

**(lxiii) Which measure is used for regression model evaluation?**

- a) Accuracy
- b) Recall
- c) Precision
- d) Mean Squared Error

**(lxiv) What does the term 'feature scaling' refer to?**

- a) Changing data types
- b) Adjusting ranges of features
- c) Encoding labels
- d) Dropping null values

**(I xv) What is the purpose of validation data?**

- a) Train model
- b) Test model
- c) Tune model
- d) Preprocess data

**(I xvi) Which process minimizes the error during training?**

- a) Dropout
- b) Activation
- c) Gradient Descent
- d) One-hot encoding

**(I xvii) What is an advantage of decision trees?**

- a) Require a lot of preprocessing
- b) Handle nonlinear data
- c) Only work with numeric data
- d) Difficult to interpret

**(I xviii) Which library is used for machine learning in Python?**

- a) NumPy
- b) Matplotlib
- c) Scikit-learn
- d) Pandas

**(I xix) Which of these is an ensemble learning method?**

- a) Logistic Regression
- b) K-NN
- c) Random Forest
- d) PCA

**(I xx) What does underfitting mean?**

- a) Model fits training data well
- b) Model generalizes well
- c) Model fails to capture patterns
- d) Model performs well on test set

**(I xxi) Which parameter is adjusted during training?**

- a) Feature
- b) Loss
- c) Weight
- d) Epoch

**(I xxii) Which is not a supervised learning algorithm?**

- a) Linear Regression
- b) K-Means
- c) Decision Tree
- d) SVM

**(I xxiii) Which method is used to reduce overfitting?**

- a) Normalization
- b) Dropout
- c) Feature selection
- d) All of the above

**(I xxiv) What is one-hot encoding used for?**

- a) Scaling data
- b) Cleaning data
- c) Encoding categorical variables
- d) Evaluating models

**(lxxv) Which of the following splits data randomly?**

- a) Standardization
- b) Train\_test\_split
- c) Label Encoding
- d) Gradient Descent

**(lxxvi) What is model overfitting?**

- a) Model fits training data too well
- b) Model performs poorly on training data
- c) Model ignores features
- d) Model always predicts one class

**(lxxvii) What is regularization in machine learning?**

- a) A technique for data cleaning
- b) Reducing model complexity to prevent overfitting
- c) Scaling feature values
- d) Increasing model depth

**(lxxviii) What does the learning rate control in training?**

- a) Model accuracy
- b) Step size during gradient descent
- c) Number of layers
- d) Number of data points

**(lxxix) Which technique helps avoid overfitting in neural networks?**

- a) Learning rate
- b) ReLU activation
- c) Dropout
- d) Cross-entropy

**(lxxx) What is the purpose of using activation functions in neural networks?**

- a) Add noise
- b) Introduce non-linearity
- c) Remove bias
- d) Normalize data

**(lxxxii) Which of the following best demonstrates how training data is used in a supervised learning pipeline?**

- a) Tuning hyperparameters after deployment
- b) Training the model using labeled examples to learn patterns
- c) Filtering out duplicate test samples
- d) Measuring bias from the model's residuals

**(lxxxii) Which one is an example of batch training?**

- a) Single record per epoch
- b) Full dataset in memory
- c) Training without target
- d) Stochastic updates

**(lxxxiii) During model development, why is splitting data into training and test sets necessary?**

- a) It prevents the need for data preprocessing
- b) It ensures the model generalizes well on unseen data
- c) It increases the number of features for the model
- d) It reduces training time by discarding half the data



**(lxxxiv) What happens if we use test data for training?**

- a) Better accuracy
- b) Fair evaluation
- c) Overfitting risk
- d) Data augmentation

**(lxxxv) While evaluating a classification model, a test set yields 90 correct predictions out of 100. What is the model's accuracy?**

- a) 0.1
- b) 0.85
- c) 0.9
- d) 1.1

**(lxxxvi) How can machine learning enhance decision-making in business operations?**

- a) By eliminating all human jobs
- b) By automating employee scheduling
- c) By randomly guessing outcomes
- d) By disabling manual processes

**(lxxxvii) Which of the following best describes a business problem suitable for machine learning?**

- a) A decision with no data involved
- b) A task based on random opinions
- c) A one-time decision by management
- d) A repetitive, data-driven decision task

**(lxxxviii) Why is it important to define a business problem clearly before applying machine learning?**

- a) To eliminate the need for data
- b) To get more funding
- c) To avoid choosing the wrong algorithm
- d) To speed up hiring

**(lxxxix) What is a primary goal of executing a machine learning pilot project in business?**

- a) To test the model in a controlled environment
- b) To launch full-scale automation
- c) To train all employees in coding
- d) To advertise the company's product

**(xc) Which step comes first when applying machine learning to solve a business problem?**

- a) Deploying the final model
- b) Understanding the business challenge
- c) Collecting customer feedback
- d) Hiring new developers

**(xci) What is the role of domain knowledge in machine learning for business?**

- a) Has no relevance
- b) Slows down automation
- c) Helps frame the problem accurately
- d) Prevents use of AI

**(xcii) Which machine learning technique is commonly used for fraud detection?**

- a) Classification
- b) Clustering
- c) Regression
- d) Reinforcement Learning

**(xciii) How does machine learning help proactively respond to IT issues?**

- a) By randomly checking logs
- b) By eliminating all network staff
- c) By predicting system failures before they occur
- d) By switching off servers

**(xciv) What is one challenge in applying machine learning to patient health records?**

- a) Excessive manual intervention
- b) Unlimited data availability
- c) No need for accuracy
- d) Data privacy concerns

**(xcv) Which type of dataset is most suitable for training a machine learning model to detect online transaction fraud?**

- a) News articles
- b) Labeled historical transaction data
- c) Product advertisements
- d) Employee surveys

**(xcvi) Why is selecting the right ML model important in solving business problems?**

- a) It guarantees faster hiring
- b) It increases website traffic
- c) It ensures accurate predictions
- d) It removes all data

**(xcvii) How does ML help improve customer service in business?**

- a) By increasing manual ticketing
- b) By enabling chatbots to provide instant responses
- c) By blocking user access
- d) By creating human-only service desks

**(xcviii) What is meant by 'supervised learning' in the business context?**

- a) Learning with labeled datasets for prediction
- b) Random decision-making
- c) Human-only feedback loops
- d) Real-time translation

**(xcix) What is the benefit of using classification models in fraud detection?**

- a) They recommend new products
- b) They count the total transactions
- c) They identify customer birthdays
- d) They separate fraudulent from legitimate transactions

**(c) What does a confusion matrix help a business evaluate?**

- a) Accuracy of classification predictions
- b) Employee count
- c) Budget estimation
- d) Customer preferences

**(ci) In healthcare, how does ML aid in diagnosis?**

- a) By encrypting lab reports
- b) By replacing all doctors
- c) By analyzing patient data to predict conditions
- d) By increasing prescription prices

**(cii) What is a feature in a machine learning model?**

- a) A code version
- b) A bug in the software
- c) A user password
- d) An input variable used to make predictions

**(ciii) Why is historical data crucial in machine learning for business?**

- a) It prevents internet access
- b) It helps the model learn from past trends
- c) It reduces programming errors
- d) It increases power consumption

**(civ) Which business metric can ML optimize in supply chain management?**

- a) Website color themes
- b) Employee bonuses
- c) Delivery time prediction
- d) Canteen menu

**(cv) What is overfitting in ML, and why is it a problem for business applications?**

- a) Model performs well on training data but poorly on new data
- b) It causes system crashes
- c) It delays product launches
- d) It improves long-term predictions

**(cvi) What is the main purpose of using ML in business problem-solving?**

- a) To increase code complexity
- b) To interpret data patterns for decision-making
- c) To replace project managers
- d) To minimize customer reviews

**(cvii) How does machine learning help reduce human bias in business decisions?**

- a) By randomly changing input data
- b) By relying solely on team opinions
- c) By using data-driven models
- d) By removing all employees

**(cviii) Why should a company monitor a machine learning model after deployment?**

- a) To keep employees engaged
- b) To maintain prediction performance over time
- c) To avoid using new data
- d) To reduce internet usage

**(cix) How can ML assist a retail store in improving inventory management?**

- a) By locking items
- b) By tracking customer age
- c) By predicting product demand trends
- d) By printing invoices

**(cx) What role does data labeling play in business-focused ML systems?**

- a) It reduces internet bills
- b) It helps the model understand input-output relationships
- c) It slows down training
- d) It deletes old files

**(cxi) Why is data quality important in a business ML project?**

- a) To increase storage size
- b) To ensure reliable model outcomes
- c) To speed up hiring
- d) To avoid using cloud services

**(cxii) What defines a successful ML project in a business environment?**

- a) Minimal user interaction
- b) Fast programming time
- c) Achieving valuable and actionable outcomes
- d) Replacing all dashboards

**(cxiii) Why is model interpretability important in business use cases?**

- a) To confuse end-users
- b) To explain decisions to stakeholders
- c) To reduce data storage
- d) To improve animations

**(cxiv) In business ML, what is the significance of using test data?**

- a) To entertain the team
- b) To evaluate the model on unseen cases
- c) To break the algorithm
- d) To make presentations better

**(cxv) How can ML support product recommendation engines?**

- a) By showing ads only
- b) By using past customer behavior data
- c) By random selection
- d) By asking for customer reviews only

**(cxvi) Which component is essential before deploying an ML solution in a business process?**

- a) Data entry form
- b) Stakeholder validation and testing
- c) Internet speed check
- d) Company branding update

**(cxvii) How can machine learning help businesses reduce customer churn?**

- a) By ignoring customer complaints
- b) By changing usernames
- c) By removing chatbots
- d) By analyzing behavior patterns and feedback

**(cxviii) Why is choosing the right features important in ML-based business models?**

- a) To speed up emails
- b) To improve model accuracy and relevance
- c) To match website colors
- d) To reduce meetings

**(cxix) What is the function of preprocessing in ML for business applications?**

- a) To clean and prepare data for learning
- b) To entertain customers
- c) To reduce data ownership
- d) To delete web pages

**(cxx) In business fraud detection, what is an example of a useful feature?**

- a) Product color
- b) Transaction time and amount
- c) Employee birthdate
- d) Company website font

**(cxxi) Why is data privacy important in healthcare ML solutions?**

- a) To reduce power bills
- b) To comply with regulations and protect patient data
- c) To slow down analytics
- d) To allow public access

**(cxxii) How does ML contribute to supply chain transparency?**

- a) By hiding vendor names
- b) By increasing paperwork
- c) By tracking goods in real time
- d) By randomizing shipping routes

**(cxxiii) What is a practical use of ML in financial services?**

- a) Predicting customer credit risk
- b) Updating logos
- c) Training new tellers
- d) Building mobile games

**(cxxiv) How does ML help HR departments in businesses?**

- a) By forecasting employee attrition
- b) By assigning manual tasks
- c) By disabling feedback systems
- d) By deleting resumes

**(cxxv) Why are regular updates needed in business ML models?**

- a) To impress customers
- b) To reflect changing data patterns and business needs
- c) To restart servers
- d) To redesign the UI

**(cxxvi) What is the role of bootstrap sampling in bagging?**

- a) To test model accuracy
- b) To generate varied datasets for training
- c) To remove outliers
- d) To clean noisy data

**(cxxvii) In boosting, what happens to the weight of a correctly classified sample?**

- a) It is set to zero
- b) It increases
- c) It decreases
- d) It is unchanged

**(cxxviii) How does stacking differ from simple voting in ensembles?**

- a) Stacking uses weights instead of probabilities
- b) Stacking uses a model to combine outputs
- c) Stacking ignores base models
- d) Stacking is used only in classification

**(cxxix) Which ensemble technique works best when base models are prone to high variance?**

- a) Stacking
- b) Boosting
- c) Bagging
- d) ECOC

**(cxxx) What type of data structure is often used to implement Random Forests?**

- a) Neural networks
- b) Decision trees
- c) K-means clusters
- d) SVMs

**(cxxxix) What is the result of combining weak learners in AdaBoost?**

- a) One strong learner
- b) Multiple random classifiers
- c) Deep learning model
- d) Reduced training time

**(cxxxix) Which ensemble method is most sensitive to noisy data?**

- a) Bagging
- b) Boosting
- c) Random Forest
- d) Voting

**(cxxxix) What is the purpose of using diverse models in ensemble learning?**

- a) To reduce runtime
- b) To avoid training
- c) To improve robustness and generalization
- d) To build fewer models

**(cxxxix) What does the final model in stacking learn from?**

- a) The training data directly
- b) The output of a random model
- c) The predictions of base models
- d) The testing errors

**(cxxxix) Which method corrects bias by giving more weight to errors in prior models?**

- a) Random Forest
- b) Bagging
- c) Boosting
- d) K-Means

**(cxxxix) What aspect of a Random Forest model reduces overfitting?**

- a) Using more hidden layers
- b) Limiting the depth
- c) Averaging across multiple trees
- d) Combining all leaves into one

**(cxxxix) How does ECOC handle multi-class classification?**

- a) Converts to single binary task
- b) Uses feature scaling
- c) Uses unsupervised learning
- d) Splits into multiple binary problems

**(cxxxix) Which of the following is true about ensemble models in general?**

- a) They often outperform individual models
- b) They reduce accuracy by averaging errors
- c) They increase model complexity for no gain
- d) They are used only with deep learning

**(cxxxix) In bagging, how are the models trained?**

- a) Sequentially on errors
- b) On the same data repeatedly
- c) On different subsets drawn with replacement
- d) Using gradient descent

**(cxl) What kind of learners does boosting focus on?**

- a) Accurate from the beginning
- b) Weak learners that can be improved
- c) Always underfit
- d) Overfit deep models

**(cxli) How does hard voting make final predictions?**

- a) Using a probability average
- b) Using the median output
- c) Using majority class prediction
- d) Using model confidence scores

**(cxlii) Which ensemble approach helps reduce model variance the most?**

- a) Bagging
- b) Boosting
- c) Stacking
- d) ECOC

**(cxliii) What happens to diversity among models in bagging?**

- a) It is eliminated
- b) It is introduced through data sampling
- c) It is introduced using feature scaling
- d) It is ignored

**(cxliv) Why is soft voting considered more flexible than hard voting?**

- a) It ignores class labels
- b) It requires fewer models
- c) It uses random forests
- d) It considers confidence/probability scores

**(cxlv) What is a potential drawback of stacking?**

- a) It uses fewer models
- b) It cannot be interpreted easily
- c) It eliminates overfitting
- d) It reduces the number of features

**(cxlvi) In ensemble learning, why is combining models useful?**

- a) It increases complexity
- b) It improves accuracy and robustness
- c) It avoids model training
- d) It speeds up testing

**(cxlvii) What is the role of base learners in boosting?**

- a) Make random predictions
- b) Correct errors made by earlier models
- c) Select features
- d) Store model outputs

**(cxlviii) Why is overfitting less likely in bagging?**

- a) It uses fixed parameters
- b) It removes all features
- c) It averages results across models
- d) It trains on biased data

**(cxlix) What makes Random Forest a type of bagging?**

- a) It uses multiple trees trained on subsets
- b) It uses random noise
- c) It builds only one large tree
- d) It uses unsupervised learning

**(cl) Why is boosting considered sequential?**

- a) It trains models on unrelated data
- b) Each model is trained after correcting the previous model's errors
- c) It uses parallel computing
- d) It avoids iteration

**(cli) What does stacking require to combine model outputs?**

- a) Dropout layers
- b) Missing value imputation
- c) Feature normalization
- d) A meta-learner

**(clii) What does ECOC convert multiclass problems into?**

- a) Neural layers
- b) Binary classification tasks
- c) Feature sets
- d) Random clusters

**(cliii) How is the final prediction made in Random Forest for classification?**

- a) Mode of tree outputs
- b) Median of results
- c) Mean of error rates
- d) First model's output

**(cliv) Why do ensemble methods generally perform better than single models?**

- a) They use less data
- b) They reduce bias and variance through aggregation
- c) They require fewer iterations
- d) They depend on a single algorithm

**(clv) What is one advantage of soft voting?**

- a) It always overfits
- b) It uses only the first model
- c) It considers prediction confidence
- d) It removes class labels

**(clvi) In stacking, what is trained on the outputs of base learners?**

- a) A loss function
- b) A final predictive model (meta-model)
- c) A noise filter
- d) A data preprocessor

**(clvii) What type of learners are most often used in bagging?**

- a) Weak learners
- b) Overfit learners
- c) Independent learners
- d) Deep networks

**(clviii) In boosting, how are difficult samples handled over iterations?**

- a) They are removed
- b) Their influence increases
- c) Their labels are flipped
- d) They are given random weights

**(clix) Which ensemble technique uses bootstrapped datasets?**



- a) AdaBoost
- c) ECOC

- b) Stacking
- d) Random Forest

**(clx) Why does stacking often use different types of models as base learners?**

- a) To reduce accuracy
- c) To maximize diversity in predictions
- b) To increase model complexity
- d) To eliminate the need for testing

### **Short Answer Type Questions**

2. What is Machine Learning?
3. Define Artificial Intelligence.
4. What is the main objective of descriptive analytics?
5. Define predictive analytics.
6. What is Big Data?
7. How is Machine Learning related to Artificial Intelligence?
8. What does "leveraging machine learning" mean?
9. What is the difference between AI and ML in one sentence?
10. Analyze the differences between descriptive and predictive analytics by comparing their objectives, methods, and outcomes using suitable examples
11. What are parametric and non-parametric model?
12. Explain the concept of supervised learning with an example
13. Explain how unsupervised learning works and give an example.
14. Differentiate-between Artificial intelligence, Machine Learning, and Deep Learning.
15. Explain an application of supervised learning and how it is used in that context.
16. Explain how supervised learning differs from unsupervised learning with an example.
17. Describe the role of data mining in machine learning with suitable examples.
18. Explain the concept of a neural network in machine learning with an example.
19. Describe any two machine learning techniques and explain how they are used.
20. Discuss the different types of Machine Learning.
21. Find and analyze the mean, median, mode, and range for the given data: 90, 94, 53, 68, 79, 94, 53, 65, 87, 90, 70, 69, 65, 89, 85, 53, 47, 61, 27, 80
22. Justify the use of reinforcement learning in dynamic environments.
23. Evaluate the effectiveness of neural networks compared to traditional models.
24. Apply one method to split a dataset into training and testing sets.
25. How would you implement a basic classification model in Python?
26. How can normalization affect model training?
27. Demonstrate how to apply feature selection to a dataset.

28. Apply any algorithm to solve a binary classification problem.
29. How would you handle missing values before training a model?
30. How can you identify overfitting in a trained model?
31. Apply cross-validation to validate model performance.
32. How would you prepare categorical data for training?
33. What step would you take to train a decision tree model in Python?
34. How can you apply scaling in your preprocessing pipeline?
35. How would you tune hyperparameters in a training model?
36. Evaluate the accuracy of a model using precision, recall, and F1-score.
37. Justify the selection of a supervised learning algorithm over an unsupervised one for a given labeled dataset.
38. Explain how machine learning supports business decision-making.
39. Explain the Difference Between Classification and Regression?
40. How does ML help in improving patient health outcomes?
41. Describe the role of ML in detecting fraud in banking systems.
42. Why is it important to focus on a business problem before selecting an ML model?
43. Explain the importance of choosing the best-fit learning model for a business case.
44. How does ML help in proactively managing IT issues?
45. How does training a model on real business data improve accuracy?
46. Why is domain knowledge important in applying ML to business problems?
47. What are the key phases of applying ML to solve business problems?
48. How does ML contribute to risk reduction in enterprises?
49. Explain how machine learning can enhance customer experience in online businesses.
50. Evaluate the performance differences between Random Forest and AdaBoost in ensemble learning.
51. Justify the use of stacking over voting in ensemble models.

Evaluate the accuracy of a model from confusion Matrix.

52.

	YES	NO
YES	16	4
NO	3	12

### Long Answer Type Questions

53. Define Machine Learning. Discuss its key principles with examples.
54. What is Artificial Intelligence? How is Machine Learning a part of AI?
55. Define Big Data. How is Big Data used in the context of Machine Learning?
56. Explain the concept of supervised and unsupervised learning with examples.
57. Describe the role of statistics and data mining in the machine learning process.
58. Explain how machine learning techniques can be applied to meet different business needs.
59. Analyze the key differences among supervised, unsupervised, and reinforcement learning approaches.
60. Analyze a real-world problem where supervised learning is best suited.
61. Analyze how neural networks differ from traditional machine learning models.
62. Compare and contrast the use of K-Means and Hierarchical Clustering for unsupervised learning tasks.
63. Analyze how reinforcement learning is applied in recommendation systems or robotics.
64. Analyze the challenges in applying machine learning techniques to real-world business data.
65. Analyze the workflow of a typical machine learning application from data collection to model evaluation.
66. Analyze a case where the wrong choice of learning method led to model inefficiency.
67. Apply the complete workflow of training a machine learning model using Python with an example.
68. Demonstrate how to preprocess a dataset and train it using at least two algorithms.
69. Show how cross-validation can be implemented and interpreted in model training.
70. Evaluate the performance of a trained model using various metrics. Which metrics are most useful and why?
71. Critically evaluate the trade-offs between bias and variance in model training.
72. Justify the need for data preprocessing before training a machine learning model.
73. Evaluate the advantages and limitations of using neural networks for training large datasets.
74. Compare and evaluate decision tree and random forest algorithms for training purposes.
75. Discuss how hyperparameter tuning affects the performance of a machine learning model.
76. Evaluate the machine learning cycle. Why is it important to iterate through the training stages?
77. Analyze the effectiveness of ensemble techniques in improving model training outcomes.
78. Explain how machine learning can be applied to solve real-world business problems.
79. Design a machine learning solution for early prediction of patient readmission in hospitals.
80. Create a pilot machine learning project to detect IT infrastructure failures before they occur.
81. Develop a machine learning model to detect fraudulent transactions in real-time.
82. Propose a solution using machine learning to enhance customer retention in a telecom business.
83. Create a model-based approach to reduce false positives in fraud detection using ensemble techniques.
84. Design a system that uses ML to recommend personalized treatment plans for patients.
85. Propose a framework where ML helps proactively manage server downtimes in cloud-based IT systems.
86. Create an ML-based approach for credit risk analysis in banking with data-driven insights.
87. Evaluate the strengths and weaknesses of bagging and boosting techniques.
88. Compare the performance of Random Forest with AdaBoost in handling noisy datasets.
89. Evaluate different ensemble techniques in terms of model interpretability and computational cost.
90. Assess the impact of hyperparameter tuning in boosting algorithms.
91. Justify the preference of ensemble models over individual models in high-stakes decision systems.

92. Evaluate the role of voting strategies in ensemble learning.
93. Critically evaluate the use of stacking classifiers in multi-class classification tasks.
94. Assess when ensemble models may not be the right choice in a business application.
95. Design an ensemble model for customer churn prediction using bagging and boosting techniques.
96. Create a hybrid ensemble model combining stacking and voting techniques.
97. Build an ensemble learning pipeline for email spam detection.
98. Design a use case where stacking improves performance over standalone models.
99. Create a comprehensive evaluation framework to compare bagging, boosting, and stacking.
100. Propose an ensemble model to predict student performance using diverse learning algorithms.
101. Design a real-time fraud detection system using ensemble methods.
102. Construct an ensemble-based medical diagnosis model to support clinical decisions.