1. ML, AI, Big Data, Descriptive Analytics, Predictive Analytics. Define, Explain characteristics, Example, Disadvantage, and advantage, Objectives, Tools (SAS, python, R).

Ans: **1. Machine Learning (ML)**

- **Definition:** ML is a branch of AI where machines learn from data and improve performance without being directly programmed.
- **Characteristics:** Learns patterns, adapts automatically, improves accuracy over time.
- **Example:** Spam email detection.
- **Advantage:** Improves efficiency, handles large datasets.
- **Disadvantage:** Needs a lot of data and computing power.
- **Objective:** Automate decision-making and predictions.
- **Tools:** Python, R, TensorFlow, Scikit-learn.

2. Artificial Intelligence (AI)

- **Definition:** AI is the simulation of human intelligence in machines to perform tasks like learning, reasoning, and problem-solving.
- **Characteristics:** Self-learning, reasoning, decision-making, problem-solving.
- **Example:** Chatbots like Siri or Alexa.
- **Advantage:** Reduces human effort, works 24/7, improves productivity.
- **Disadvantage:** Expensive, may replace human jobs.
- **Objective:** Create intelligent systems that think and act like humans.
- **Tools:** Python, R, TensorFlow, SAS.

3. Big Data

- **Definition:** Big Data refers to extremely large datasets that are difficult to process using traditional methods.
- **Characteristics:** 5Vs – Volume, Velocity, Variety, Veracity, Value.
- **Example:** Social media data like tweets and posts.
- **Advantage:** Provides better decision-making, discovers hidden patterns.
- **Disadvantage:** Requires high storage, costly tools, and privacy issues.
- **Objective:** Extract useful insights from massive data.
- **Tools:** Hadoop, Spark, Python, R.

4. Descriptive Analytics

- **Definition:** Descriptive analytics explains past data and answers “What happened?”.
- **Characteristics:** Simple, historical, fact-based.
- **Example:** Sales reports showing last month’s revenue.
- **Advantage:** Easy to understand, helps identify trends.
- **Disadvantage:** Only looks at the past, no future prediction.
- **Objective:** Summarize past events for better understanding.
- **Tools:** Excel, SAS, Python, R.

5. Predictive Analytics

- **Definition:** Predictive analytics uses past data and models to predict future outcomes.
- **Characteristics:** Data-driven, uses ML models, probability-based.
- **Example:** Predicting customer churn in telecom.
- **Advantage:** Improves planning and reduces risks.
- **Disadvantage:** Not 100% accurate, depends on data quality.
- **Objective:** Forecast future trends and outcomes.
- **Tools:** R, Python, SAS, RapidMiner.

SAS

- **Definition:** SAS (Statistical Analysis System) is a software suite used for data management, advanced analytics, and reporting.
- **Features:** Easy for statistical analysis, reliable, widely used in industries.
- **Limitation:** Expensive license, less flexible than open-source.

Python

- **Definition:** Python is an open-source programming language used in data science, ML, and AI.
- **Features:** Simple syntax, large libraries (NumPy, Pandas, TensorFlow).
- **Limitation:** Slower execution compared to some compiled languages.

R

- **Definition:** R is an open-source language mainly used for statistics and data visualization.
- **Features:** Strong in graphs, supports advanced statistical models.
- **Limitation:** Slower with very large datasets, less general-purpose.

1. Difference between ML and AI (5 points)

Point	Artificial Intelligence (AI)	Machine Learning (ML)
Definition	Field of creating intelligent machines	Subset of AI focused on learning from data
Goal	Achieve human-like reasoning & decision making	Predict outcomes & improve with experience
Approach	Uses rules, reasoning, problem-solving	Uses data, algorithms, models
Example	Self-driving car, Chatbot	Spam filter, Recommendation system
Scope	Broad field	Narrower subfield

2. Difference between Descriptive vs Predictive Analytics (5 points)

Point	Descriptive Analytics	Predictive Analytics
Definition	Explains past events (“What happened?”)	Predicts future events (“What may happen?”)
Data Used	Historical data only	Past + present data
Output	Reports, summaries, dashboards	Forecasts, probabilities
Purpose	Understand past performance	Improve planning & reduce risks
Example	Last month’s sales report	Predicting customer churn

3. Principles of ML

Ans:

- Learn from data automatically.
- Improve performance with more experience.
- Use training and testing datasets.
- Find hidden patterns & relationships.
- Automate predictions and decisions.

4. Big Data in context with ML

Ans:

- ML needs huge datasets to train models.
- Big Data provides volume, velocity, and variety.
- Together they improve prediction accuracy.
- Example: Netflix recommendations use Big Data + ML.
- Without Big Data, ML cannot scale well.

5. Leveraging the power of ML (2/5 marks)

Ans:

- Using ML to analyze data automatically.
- Helps faster and smarter decisions.
- Reduces human errors and saves cost.
- Applications: fraud detection, healthcare, e-commerce.
- Makes systems adaptive & intelligent.

6. Relation between AI and Big Data / Big Data in ML (with example)

Ans:

- AI needs Big Data for learning & reasoning.
- Big Data supplies structured and unstructured datasets.
- Example: Google AI uses search data for better results.
- ML uses Big Data for training accurate models.
- Together they create real-world intelligent apps.

7. Context of ML, AI and their types

Ans:

- **AI:** Broad field of creating intelligent systems.
- **ML:** Subset of AI that learns from data.
- **Types of ML:**
  - Supervised → Uses labeled data.
  - Unsupervised → Finds hidden patterns.
  - Reinforcement → Learns via rewards/punishments.

8. 5Vs of Big Data (Volume, Velocity, Variety, Veracity, Value)

Ans:

- **Volume:** Huge data size (TB, PB).
- **Velocity:** Fast data generation (real-time).
- **Variety:** Different forms (text, video, images).
- **Veracity:** Accuracy & reliability of data.
- **Value:** Business insights gained.

9. Big Data Learning – Definition

Ans:

- Using ML/AI to analyze very large datasets.
- Identifies hidden patterns & predictions.
- Example: Analyzing billions of health records for disease prediction.

10. Types of ML

Ans:

- **Supervised Learning:** With labeled data (spam detection).
- **Unsupervised Learning:** No labels, finds groups (customer clustering).
- **Reinforcement Learning:** Learns via feedback (game bots).

11. ML vs Traditional Computing (5 points)

Point	Traditional Computing	Machine Learning
Rules	Fixed rules coded by humans	Learns rules automatically from data
Adaptability	Cannot adapt to new data	Improves with more data
Data Use	Limited	Data-driven
Example	Calculator	Speech recognition
Improvement	No self-improvement	Continuous improvement

Q1. What is the primary role of Predictive Analytics?
a) Summarizing past data
b) Forecasting future outcomes ✅

c) Cleaning raw data
d) Visualizing reports

Q2. Which technique is mainly used in Descriptive Analytics?
a) Regression
b) Clustering
c) Data Summarization and Visualization ✅
d) Neural Networks

Q3. Which of the following is an application example of Machine Learning?
a) Weather prediction ✅
b) File storage
c) Basic calculators
d) Manual data entry

Q4. Which are subfields of Artificial Intelligence (AI)?
a) Machine Learning ✅
b) Deep Learning ✅
c) Natural Language Processing ✅
d) All of the above ✅

(Answer: d)

Q5. Which of the following is a limitation of Machine Learning?
a) Requires large amounts of data ✅
b) Fully eliminates human intervention
c) No computational cost
d) Always provides 100% accuracy

Q6. Which tools are commonly used in Predictive Analytics?
a) Python ✅
b) R ✅
c) SAS ✅
d) All of the above ✅

(Answer: d)

Q7. Which of the following are branches of Machine Learning?
a) Supervised Learning ✅
b) Unsupervised Learning ✅
c) Reinforcement Learning ✅
d) All of the above ✅

(Answer: d)

Q8. Machine Learning vs Traditional Computing – the main difference is:
a) ML learns from data and improves automatically ✅
b) Traditional computing always requires explicit programming ✅
c) Both a and b ✅

(Answer: c)

Q9. In ML, "Learning" means:
a) Feeding data into memory
b) System improving performance from experience/data ✅
c) Writing new programs manually
d) Using more storage space

1. Use of Training Data and Testing Data in ML

Ans:

- **Training Data:** Dataset used to train the ML model so it learns patterns and relationships.
- **Testing Data:** Separate dataset to evaluate how well the trained model performs on unseen data.
- Ensures model accuracy and avoids overfitting.
- Example: Train a spam filter with labeled emails, test on new emails.

2. Types of Machine Learning – Definition and Example

Ans:

Type	Definition	Example
Supervised Learning	Learns from labeled data with input-output mapping	Spam email detection, House price prediction
Unsupervised Learning	Learns from unlabeled data, identifies patterns	Customer segmentation, Market basket analysis
Reinforcement Learning	Learns through feedback (rewards & punishment)	Self-driving cars, Game-playing AI (Chess, Go)

3. Example of all ML Types

Ans:

- **Supervised:** Predicting stock prices based on past data.
- **Unsupervised:** Clustering customers by purchase behavior.

- **Reinforcement:** AI learns to play Atari games by maximizing score.

4. Type of Learning that Helps Identify Patterns in Data

Ans: Unsupervised Learning ✅

- Learns structure, groups, or clusters without pre-labeled outcomes.

5. Statistics – Mean, Median, Mode

Ans:

- **Mean:** Sum of all values ÷ number of values.
- **Median:** Middle value in sorted dataset.
- **Mode:** Most frequent value in the dataset.
- Example: For [2, 3, 3, 6], Mean=3.5, Median=3, Mode=3

6. Predictive Analytics

Ans:

- Uses historical data and ML/statistical models to forecast future events.
- Helps in planning, decision making, and risk reduction.
- Example: Predicting telecom customer churn, future sales, or disease outbreaks.

7. Difference Between Supervised and Unsupervised Learning

Feature	Supervised Learning	Unsupervised Learning
Data	Labeled	Unlabeled
Goal	Predict outputs	Discover patterns
Feedback	Uses error correction	No feedback
Example	Email spam detection	Customer clustering
Algorithm	Regression, Classification	K-means, Hierarchical clustering

8. ML Type Used for Rewards and Punishments

Ans: Reinforcement Learning ✅

- Learns through trial and error.

- Maximizes rewards and minimizes penalties.
- Example: Game bots, self-driving cars navigation.

9. Neural Network – Mimics Human Brain

Ans:

- ML model with layers of neurons: Input, Hidden, Output.
- Each neuron processes data, passes signal forward.
- Learns patterns by adjusting weights.
- Example: Handwriting recognition, speech recognition.

1. Applications of ML – Image, Classification, Email/Spam, Healthcare

Ans:

- **Image:** ML algorithms, especially convolutional neural networks (CNNs), can detect and classify objects in images.
  - **Example:** Facial recognition in security systems, detecting tumors in X-ray or MRI scans.
- **Classification:** ML is used to categorize data into classes based on features.
  - **Example:** Classifying emails as spam or not spam, predicting whether a customer will churn.
- **Email/Spam Detection:** Supervised ML models like Naive Bayes or Decision Trees can filter unwanted emails automatically.
  - **Benefit:** Saves time, reduces human effort, improves productivity.
- **Healthcare:** ML predicts diseases, recommends treatments, or analyzes medical images.
  - **Example:** Predicting diabetes risk from patient data, detecting cancer cells from pathology images.

2. Which Helps in Decision Making for a Business Operation?

Question:

a. randomly guessing outcomes
b. eliminating all human jobs
c. disabling manual process
d. automating employee schedule

Ans: d. Automating employee schedule ✅

- **Explanation:**
  - Automation reduces errors, speeds up repetitive tasks, and improves efficiency.
  - ML and predictive analytics can optimize workforce allocation, ensuring the right employees are scheduled at the right time.
  - Random guessing or disabling manual processes without logic cannot aid decision-making effectively.

3. Objective to Apply ML in a Business Problem

Question:

a. To get more of finding
b. To choose the correct algorithm
c. To eliminate need for any data
d. To speed HR operation

Ans: b. To choose the correct algorithm ✅

- **Explanation:**
  - Applying ML effectively requires selecting the appropriate algorithm that fits the data type, business objective, and problem complexity.
  - Correct algorithm choice ensures better predictions, insights, and automation.
  - The goal is not merely speed or ignoring data—it is to model data accurately for informed business decisions.

4. Algorithm Applicable for Fraud Detection

Question:

a. Regression
b. Classification
c. Clustering
d. Neural

Ans: b. Classification ✅

- **Explanation:**
  - Fraud detection is a **binary classification problem**: transactions are either fraudulent or legitimate.
  - Supervised learning algorithms like **Decision Trees, Random Forest, and Logistic Regression** are commonly used.
  - Clustering (unsupervised) can detect anomalies, but primary detection uses classification.
  - Neural networks can also be used for complex fraud detection, but the simplest and most direct is classification.

1. Role of Data Mining in Machine Learning

Ans:

- **Definition:** Data mining is the process of discovering patterns, correlations, trends, and insights from large datasets.
- **Role in ML:**
  - Provides meaningful features for model training.
  - Helps identify hidden relationships in data for predictive modeling.
  - Reduces data dimensionality for faster, more accurate ML algorithms.
- **Example:**

- In retail, data mining uncovers frequent purchase patterns which are then used to build recommendation systems using ML.
- In banking, patterns of suspicious transactions can be mined to train fraud detection models.

2. Justification for Using Supervised Algorithm on Labeled Dataset

Ans:

- **Definition:** Supervised learning uses labeled datasets where input-output pairs are known.
- **Reason for Selection:**
  - Labeled data allows the algorithm to learn exact mapping from features to outcomes.
  - Provides measurable feedback (error/loss) for model training.
  - Guarantees higher accuracy for prediction tasks compared to unsupervised learning.
- **Example:**
  - Predicting credit card fraud: Each transaction is labeled as “fraudulent” or “legitimate,” allowing supervised algorithms like Decision Trees or Logistic Regression to classify new transactions accurately.

3. Difference Between Classification and Regression

Feature	Classification	Regression
Output Type	Discrete (categories)	Continuous (numeric)
Goal	Assign a class/label	Predict a numerical value
Example	Email spam detection, Loan approval	House price prediction, Stock forecasting
Algorithm Examples	Decision Tree, SVM, Random Forest	Linear Regression, Polynomial Regression
Evaluation Metric	Accuracy, Precision, Recall	MSE, RMSE, R^2

- **Extra Note:** Classification answers “Which category?” while regression answers “How much?”

1. Detect Fraud in Banking System

Ans:

- **Application:** Machine Learning is widely used to detect fraudulent transactions in banking and finance.
- **How it Works:**
  - Historical transaction data (labeled as “fraud” or “legitimate”) is used to train supervised ML models.
  - Models like **Decision Trees, Random Forests, Logistic Regression, and Neural Networks** classify transactions in real-time.
- **Benefits:**
  - Detects unusual patterns or anomalies automatically.
  - Reduces financial losses due to fraud.
  - Works in real-time, faster than manual detection.
- **Example:** Credit card companies flag suspicious transactions instantly based on ML predictions.

2. Improving Patient Health Outcomes

Ans:

- **Application:** ML helps healthcare providers diagnose diseases, recommend treatments, and predict patient risks.
- **How it Works:**
  - Models analyze patient data including medical history, lab results, imaging data.
  - Supervised learning predicts disease probability; unsupervised learning finds hidden patterns in symptoms.
- **Benefits:**
  - Early disease detection (e.g., cancer, diabetes).
  - Personalized treatment plans for better recovery.
  - Reduces errors compared to traditional diagnostic methods.
- **Example:** Predicting likelihood of heart attacks using patient vitals and lifestyle data.

3. Supporting Decision-Making in Business Models

Ans:

- **Application:** ML assists businesses in making informed decisions by analyzing large datasets and predicting outcomes.
- **How it Works:**
  - Historical sales, customer behavior, and market trends are analyzed using predictive models.
  - ML recommends actions like inventory planning, pricing, marketing campaigns, or customer retention strategies.
- **Benefits:**
  - Reduces guesswork and human bias in decision-making.
  - Optimizes resource allocation and increases profitability.

    - Identifies market trends and emerging opportunities.
- **Example:** E-commerce companies use ML to recommend products to customers, forecast demand, and optimize stock levels.