Q1: What is Machine learning?
A: Machine learning is a way for computers to learn patterns from data and make predictions or decisions without being directly programmed for each rule. The model improves as it sees more examples.
Example: An email app learns from past “spam/not spam” labels to automatically filter new emails.

Q2: Mention the difference between Data Mining and Machine learning?
A: Data mining focuses on discovering hidden patterns and insights in large datasets. Machine learning focuses on building models that can predict outcomes on new data.
Example: Data mining might find that customers buy chips with soda; machine learning uses this to predict what a new customer is likely to buy.

Q3: What is ‘Overfitting’ in Machine learning?
A: Overfitting happens when a model learns the training data too well, including noise and random quirks, so it performs poorly on new data.
Example: A student memorizes answers to practice questions but fails on the real exam with new questions.

Q4: Why overfitting happens?
A: It happens when the model is too complex for the amount/quality of data, or when training goes too long, or when there’s not enough regularization.
Example: A deep tree with many branches perfectly fits training points but guesses badly on unseen points.

Q5: How can you avoid overfitting?
A: Use more data, simplify the model, apply regularization (L1/L2/dropout), stop training early, and use cross-validation.
Example: Limit a decision tree’s depth and tune it with cross-validation so it generalizes better.

Q6: What is inductive machine learning?
A: It's learning general rules from specific examples. The model "induces" a pattern from data and uses it to predict the future.
Example: From many cat images, a model learns features of cats and identifies a new cat photo.

Q7: What are the five popular algorithms of Machine Learning?
A: Linear/Logistic Regression, Decision Trees/Random Forests, Support Vector Machines, k-Nearest Neighbors, and Neural Networks.
Example: Logistic regression predicts if a transaction is fraud or not.

Q8: What are the different Algorithm techniques in Machine Learning?
A: Supervised, Unsupervised, Semi-supervised, and Reinforcement Learning; plus paradigms like ensemble methods and deep learning.
Example: Unsupervised clustering groups customers by behavior without labels.

Q9: What are the three stages to build the hypotheses or model in machine learning?
A: (1) Choose/define the model and features, (2) Train the model on data, (3) Evaluate and tune, then deploy.
Example: Pick logistic regression, train on past churn data, tune thresholds, then use it in production.

Q10: What is the standard approach to supervised learning?
A: Split data into train/validation/test, pick features, train a model, tune hyperparameters, evaluate, and deploy.
Example: Train a spam classifier on labeled emails, validate to choose parameters, test on a holdout set.

Q11: What is 'Training set' and 'Test set'?
A: The training set is used to fit the model. The test set is kept separate to check

how well the model works on new, unseen data.
Example: Train on 80% of images, test accuracy on the remaining 20%.

Q12: List down various approaches for machine learning?
A: Supervised, Unsupervised, Semi-supervised, Reinforcement Learning, and Self-supervised (common in deep learning).
Example: Reinforcement learning teaches a robot to walk by rewarding stable steps.

Q13: What is not Machine Learning?
A: Hard-coded rules or simple look-ups without learning from data aren't ML.
Example: "If user age < 18, show page A" is a fixed rule, not learning.

Q14: Explain what is the function of 'Unsupervised Learning'?
A: It finds structure in unlabeled data—like groups (clustering) or lower-dimensional representations.
Example: Group news articles by topic without any topic labels.

Q15: Explain what is the function of 'Supervised Learning'?
A: It maps inputs to known outputs using labeled data to predict future labels.
Example: Predict house prices from features like size and location using past sales.

Q16: What is algorithm independent machine learning?
A: It refers to methods or ideas that work across many algorithms, like cross-validation, feature scaling, or model selection.
Example: Using k-fold cross-validation applies to trees, SVMs, and neural nets alike.

Q17: What is the difference between artificial learning and machine learning?
A: "Artificial learning" isn't a standard term; usually we mean "Artificial Intelligence" (AI) vs ML. AI is the broader goal of making machines smart; ML is

the data-driven method to achieve it.
Example: A chess program (AI) uses ML to learn good moves from past games.

Q18: What is classifier in machine learning?
A: A classifier predicts a category (label) for an input.
Example: Classifying whether a review is “positive” or “negative.”

Q19: What are the advantages of Naive Bayes?
A: Simple, fast, works well with high-dimensional text, needs little data, and handles noisy features.
Example: Quick and accurate spam filtering on millions of emails.

Q20: What is Inductive Logic Programming in Machine Learning?
A: It learns logical rules (in first-order logic) from examples and background knowledge.
Example: Learning rules like “If X is a parent of Y and Y is a parent of Z, then X is a grandparent of Z.”

Q21: What is Model Selection in Machine Learning?
A: Choosing the best model type and hyperparameters based on validation data to balance bias and variance.
Example: Comparing random forest vs. SVM on validation accuracy and picking the winner.

Q22: What are the two methods used for the calibration in Supervised Learning?
A: Platt Scaling and Isotonic Regression are common methods to calibrate predicted probabilities.
Example: Adjusting a credit default model so a predicted 0.7 truly means ~70% chance.

Q23: Which method is frequently used to prevent overfitting?
A: Regularization (L1/L2), early stopping, and cross-validation.
Example: Adding L2 penalty in linear regression to avoid huge, unstable weights.

Q24: Why instance based learning algorithm sometimes referred as Lazy learning algorithm?
A: Because it delays learning until prediction time; it stores data and computes results when needed (e.g., kNN).
Example: kNN finds the closest past customers only when predicting for a new customer.

Q25: What are the two classification methods that SVM can handle?
A: SVM handles binary classification natively; for multi-class, use one-vs-rest or one-vs-one strategies.
Example: Building 10 one-vs-rest SVMs to classify digits 0–9.

Q26: What is ensemble learning?
A: Combining multiple models to get better performance than any single model.
Example: Voting across many decision trees (random forest) to predict churn.

Q27: When to use ensemble learning?
A: When single models are unstable or weak, when variance or bias is high, or when you need robust accuracy.
Example: Use an ensemble during a Kaggle competition to squeeze extra accuracy.

Q28: What are the two paradigms of ensemble methods?
A: Bagging (parallel, variance reduction) and Boosting (sequential, bias reduction).
Example: Random Forest is bagging; XGBoost is boosting.

Q29: What is the general principle of an ensemble method and what is bagging and boosting in ensemble method?

A: Principle: combine diverse models to reduce errors. Bagging trains models on bootstrapped samples to reduce variance; Boosting trains models sequentially to fix errors and reduce bias.
Example: Random Forest (bagging) vs. AdaBoost/XGBoost (boosting).

Q30: What is bias-variance decomposition of classification error in ensemble method?
A: Error = Bias^2 (underfitting) + Variance (overfitting) + Irreducible noise. Ensembles try to reduce variance (bagging) or bias (boosting).
Example: Random forest lowers variance by averaging many trees.

Q31: What is an Incremental Learning algorithm in ensemble?
A: It updates the model as new data arrives without retraining from scratch (online learning).
Example: A click-through-rate model updating hourly as new user clicks stream in.

Q32: What is PCA, KPCA and ICA used for?
A: PCA reduces dimensions linearly; KPCA does nonlinear reduction using kernels; ICA separates mixed signals into independent sources.
Example: ICA separates individual speakers' voices from a mixed audio recording.

Q33: What is dimension reduction in Machine Learning?
A: Reducing the number of features while keeping most of the important information.
Example: Summarize 100 sensor readings into 5 principal components for faster modeling.

Q34: What are support vector machines?
A: SVMs find the best boundary (hyperplane) that maximizes the margin between classes; with kernels, they handle complex boundaries.
Example: Classify images as cat vs. dog with an RBF-kernel SVM.

Q35: Differentiate between inductive learning and deductive learning?
A: Inductive learns general rules from examples (data → rule). Deductive applies general rules to decide outcomes (rule → case).
Example: Inductive: learn spam rules from emails; Deductive: apply known spam rules to new emails.

Q36: What is the difference between Data Mining and Machine Learning?
A: Data mining = finding patterns/insights; ML = building predictive models that generalize.
Example: Mining finds "people buy diapers with beer"; ML predicts if a shopper will buy both.

Q37: Differentiate supervised and unsupervised machine learning.
A: Supervised uses labeled data to predict labels; unsupervised uses unlabeled data to discover structure.
Example: Predict loan default (supervised) vs. cluster customers by behavior (unsupervised).

Q38: How does Machine Learning differ from Deep Learning?
A: Deep learning is a subset of ML that uses multi-layer neural networks and large data to learn complex patterns automatically.
Example: Deep learning recognizes faces in photos with high accuracy.

Q39: How is KNN different from k-means?
A: kNN is a supervised classifier using neighbors to label a point; k-means is unsupervised clustering that groups points into k clusters.
Example: kNN labels a new email as spam; k-means groups customers into segments.

Q40: What are the different types of Algorithm methods in Machine Learning?
A: Supervised, Unsupervised, Semi-supervised, Reinforcement, and Self-

supervised.
Example: A self-driving car uses reinforcement learning to improve driving policy.

Q41: What do you understand by Reinforcement Learning technique?
A: An agent learns by taking actions in an environment to maximize rewards over time.
Example: A game bot learns to win by earning points for good moves.

Q42: What is the trade-off between bias and variance?
A: Lower bias often increases variance and vice versa. The goal is the sweet spot that minimizes total error.
Example: A simple model underfits (high bias); a very complex model overfits (high variance).

Q43: How do classification and regression differ?
A: Classification predicts categories; regression predicts continuous numbers.
Example: “Spam or not” (classification) vs. “price of a car” (regression).

Q44: What are the three stages of building the hypotheses or model in machine learning?
A: Define the problem and features, train the model, evaluate and improve (same essence as Q9).
Example: For churn: choose features, train gradient boosting, tune and deploy.

Q45: Describe 'Training set' and 'training Test'.
A: Training set fits the model; the test set evaluates final performance on unseen data. (Likely meant “test set”.)
Example: Train on January–June data, test on July data.

Q46: What are the common ways to handle missing data in a dataset?
A: Drop rows/columns, impute with mean/median/mode, predictive imputation,

or “missing” category/indicator.
Example: Replace missing ages with median age and add a “was_missing_age” flag.

Q47: What are the necessary steps involved in Machine Learning Project?
A: Define problem, collect/clean data, EDA, feature engineering, model selection, training, validation, testing, deployment, and monitoring.
Example: Build a demand forecast model and monitor weekly accuracy drift.

Q48: Describe Precision and Recall?
A: Precision = of predicted positives, how many are correct. Recall = of actual positives, how many did we catch.
Example: For disease tests: precision avoids false alarms; recall ensures we catch sick patients.

Q49: What do you understand by Decision Tree in Machine Learning?
A: A tree that splits data on features to make decisions, ending in predictions at leaves.
Example: Split by “income > X?” then “age > Y?” to decide loan approval.

Q50: What do you understand by algorithm independent machine learning?
A: Cross-model practices like cross-validation, feature scaling, and ensembling that apply to many algorithms.
Example: Standardizing features helps SVM, logistic regression, and neural nets.

Q51: Describe the classifier in machine learning.
A: A classifier predicts labels based on learned patterns.
Example: A face unlock system classifies “owner” vs. “not owner.”

Q52: What is SVM in machine learning? What are the classification methods that SVM can handle?

A: SVM finds a maximum-margin boundary; with kernels it handles nonlinear data. It natively handles binary classes; multi-class via one-vs-rest or one-vs-one.
Example: Handwritten digit recognition with one-vs-one SVMs.

Q53: What do you understand by the Confusion Matrix?
A: A table showing counts of predicted vs actual classes: TP, FP, FN, TN. It summarizes classification performance.
Example: See how many spam emails were correctly/incorrectly flagged.

Q54: Explain True Positive, True Negative, False Positive, and False Negative in Confusion Matrix with an example.
A: TP: predicted positive & actually positive; TN: predicted negative & actually negative; FP: predicted positive but actually negative; FN: predicted negative but actually positive.
Example: Cancer test: FP = healthy person flagged as sick; FN = sick person missed.

Q55: What according to you, is more important between model accuracy and model performance?
A: "Performance" is broader—includes accuracy plus precision/recall/F1/AUC, speed, fairness, and cost. Prefer overall performance aligned with business goal.
Example: For fraud, high recall (catch fraud) may be more important than raw accuracy.

Q56: What is Bagging and Boosting?
A: Bagging builds many models on bootstrapped samples and averages them; boosting builds models sequentially to fix prior errors.
Example: Random Forest (bagging) vs. XGBoost (boosting).

Q57: What are the similarities and differences between bagging and boosting in Machine Learning?
A: Both combine many learners. Bagging reduces variance (parallel, independent);

Boosting reduces bias (sequential, weighted).
Example: Bagging stabilizes noisy trees; boosting lifts a weak learner's accuracy.

Q58: What do you understand by Cluster Sampling?
A: A sampling method where the population is divided into clusters, some clusters are chosen, and all or some members from selected clusters are sampled.
Example: Survey a few randomly chosen schools and test all students there.

Q59: What do you understand by the F1 score?
A: Harmonic mean of precision and recall; balances both in a single number.
Example: Used in spam detection when both catching spam and avoiding false alarms matter.

Q60: How is a decision tree pruned?
A: Limit depth/min samples, cost-complexity pruning (remove weak branches), or post-prune using validation scores.
Example: Cut branches that don't improve validation accuracy.

Q61: What are the Recommended Systems?
A: Systems that suggest items users may like using user behavior and item data (content-based, collaborative, hybrid).
Example: Netflix suggesting movies based on your watch history.

Q62: When does regularization become necessary in Machine Learning?
A: When models overfit or when features are many/noisy.
Example: L2 regularization keeps linear regression weights small and stable.

Q63: What is Regularization? What kind of problems does regularization solve?
A: Adding a penalty to model complexity to discourage overfitting. It reduces variance and improves generalization.
Example: L1 (Lasso) drives some weights to zero, simplifying the model.

Q64: Why do we need to convert categorical variables into factor? Which functions are used to perform the conversion?
A: Models need numbers; factors/encodings turn categories into numeric form (one-hot, label encoding, target encoding).
Example: Convert “Red/Blue/Green” into three binary columns for a linear model.

Q65: Do you think that treating a categorical variable as a continuous variable would result in a better predictive model?
A: Usually no; it creates fake order/spacing. Proper encoding is better unless the category truly has order.
Example: Treating “city” as numbers 1,2,3 misleads the model.

Q66: How is machine learning used in day-to-day life?
A: In recommendations, navigation, spam filters, credit scoring, speech assistants, and camera photo enhancements.
Example: Maps predicting fastest routes using learned traffic patterns.

Q67: How Do You Handle Missing or Corrupted Data in a Dataset?
A: Identify patterns, drop or impute, add missing indicators, and validate impact on models.
Example: Impute missing incomes with median and mark a “missing_income” flag.

Q68: How Can You Choose a Classifier Based on a Training Set Data Size?
A: Small data: simple models (Naive Bayes, logistic). Medium: trees, SVMs. Large/high-dim: boosted trees or neural nets. Always validate.
Example: With 1,000 rows, start with logistic regression before deep nets.

Q69: What Are the Applications of Supervised Machine Learning in Modern Businesses?
A: Demand forecasting, churn prediction, fraud detection, personalization, quality

control, and risk scoring.
Example: Bank flags suspicious transactions in real time.

Q70: What is Semi-supervised Machine Learning?
A: Learning from a small set of labeled data plus a large set of unlabeled data.
Example: Few labeled images and many unlabeled ones to train a better image classifier.

Q71: Compare K-means and KNN Algorithms.
A: k-means: unsupervised, clusters data into k groups. kNN: supervised, classifies using nearest labeled points.
Example: k-means segments customers; kNN labels a new customer's segment.

Q72: What Is 'naive' in the Naive Bayes Classifier?
A: It assumes features are independent given the class—an over-simple ("naive") assumption that often works well.
Example: Words in an email are treated as independent for spam detection.

Q73: How Will You Know Which Machine Learning Algorithm to Choose for Your Classification Problem?
A: Consider data size, feature types, linear vs. nonlinear patterns, interpretability, training time, and validate with experiments.
Example: Try logistic (baseline), tree-based (nonlinear), and SVM; pick the best on validation.

Q74: How is Amazon Able to Recommend Other Things to Buy? How Does the Recommendation Engine Work?
A: Collaborative filtering (people like you liked X), content-based (items similar to what you viewed), or hybrids.
Example: You buy a phone; it recommends cases bought by similar shoppers.

Q75: When Will You Use Classification over Regression?
A: When the output is a category, not a number.
Example: Predicting "loan approved or not" (classification) vs. "loan amount" (regression).

Q76: How Do You Design an Email Spam Filter?
A: Collect labeled emails, clean text, extract features (words/embeddings), train a classifier (Naive Bayes/logistic), tune threshold, monitor.
Example: Model flags spam and sends it to the spam folder with high precision.

Q77: What is a Random Forest?
A: An ensemble of decision trees built on bootstrapped samples with random feature selection; predictions are averaged/voted.
Example: Predicting customer churn with hundreds of diverse trees.

Q78: What is Pruning in Decision Trees, and How Is It Done?
A: Removing weak branches to avoid overfitting by limiting depth or using cost-complexity pruning with validation.
Example: Cut off branches that only improve training accuracy slightly.

Q79: Briefly Explain Logistic Regression.
A: A linear model that predicts the probability of a binary class using a sigmoid function.
Example: Predict if a user will click an ad (yes/no) from their features.

Q80: Explain the K Nearest Neighbor Algorithm.
A: To classify, find the k closest labeled points and vote; to regress, average their values.
Example: A new movie is labeled "comedy" because its 5 nearest neighbors are comedies.

Q81: What is Kernel SVM?
A: An SVM that uses a kernel function to map inputs into a higher-dimensional space to separate classes nonlinearly.
Example: Use RBF kernel to separate spiraled data that is inseparable linearly.

Q82: What Are Some Methods of Reducing Dimensionality?
A: PCA, autoencoders, feature selection (filter/wrapper), t-SNE/UMAP for visualization, and removing collinear features.
Example: Use PCA to cut 500 features down to 50 for faster training.

Q83: What is Principal Component Analysis?
A: A technique that transforms features into new uncorrelated components ordered by how much variance they capture.
Example: Compress image data to few components while keeping most detail.

Q84: What do you understand by Type I vs Type II error?
A: Type I (false positive): you say positive when it's not. Type II (false negative): you miss a real positive.
Example: Flagging a healthy person as sick (Type I) vs. missing a sick person (Type II).

Q85: Explain Correlation and Covariance.
A: Covariance shows how two variables change together (scale-dependent). Correlation is the scaled version (between −1 and 1) showing strength/direction.
Example: Height and weight have positive correlation.

Q86: What are Support Vectors in SVM?
A: Data points closest to the decision boundary that define the margin; they determine the hyperplane.
Example: Moving a support vector can change the SVM boundary.

Q87: What is Cross-Validation?
A: A way to estimate model performance by splitting data into k folds, training on k−1 folds, and testing on the remaining fold, repeating k times.
Example: 5-fold CV gives a stable estimate of accuracy before deployment.

Q88: What are the different methods to split a tree in a decision tree algorithm?
A: For classification: Gini impurity or Information Gain (entropy). For regression: variance or mean squared error reduction.
Example: Choose the feature split that most reduces Gini impurity.

Q89: How does the Support Vector Machine algorithm handle self-learning?
A: Classic SVMs are supervised; for semi-supervised "self-learning," variants like transductive SVMs use unlabeled data to refine the boundary.
Example: Use a semi-supervised SVM to leverage many unlabeled texts plus a few labeled ones.

Q90: What are the assumptions you need to take before starting with linear regression?
A: Linear relationship, independent errors, homoscedasticity (constant variance), normally distributed errors (for inference), and low multicollinearity.
Example: Plot residuals vs. predictions to check constant variance.

Q91: What is the difference between Lasso and Ridge regression?
A: Ridge (L2) shrinks weights but keeps them nonzero; Lasso (L1) can set some weights to zero (feature selection).
Example: Use Lasso to automatically drop unhelpful features.

Q92: What is Entropy in Machine Learning?
A: A measure of impurity or uncertainty in data; higher entropy means more mixed classes.
Example: A node with 50% spam and 50% not spam has high entropy.

Q93: What is Epoch in Machine Learning?
A: One complete pass of the training dataset through the learning algorithm.
Example: Training for 10 epochs means the network saw all training samples 10 times.

Q94: Differentiate between Classification and Regression in Machine Learning
A: Classification predicts class labels; regression predicts continuous values.
Example: Predict if a customer will churn (classification) vs. predict their monthly spend (regression).

Q95: How is the suitability of a Machine Learning Algorithm determined for a particular problem?
A: Match problem type, data size, feature types, need for interpretability, training time, and evaluation results on validation data.
Example: For tabular data with mixed types, try tree-based methods first.

Q96: What is ROC Curve and what does it represent?
A: A plot of True Positive Rate vs. False Positive Rate across thresholds; AUC summarizes overall separability.
Example: AUC close to 1.0 means your classifier separates classes well.

Q97: Both being Tree-based Algorithms, how is Random Forest different from Gradient Boosting Machine (GBM)?
A: Random Forest averages many independent trees (reduces variance). GBM builds trees sequentially to fix errors (reduces bias). GBM often needs more tuning.
Example: RF is robust out-of-the-box; GBM can win competitions with careful tuning.

Q98: What do you understand about the P-value?
A: The p-value is the probability of seeing results at least as extreme as yours if

the null hypothesis were true. Small p suggests evidence against the null.
Example: p = 0.01 suggests a real link between ad spend and sales.

Q99: Suppose you found that your model is suffering from high variance. Which algorithm do you think could handle this situation and why?
A: Use algorithms that reduce variance: Random Forests (bagging), Regularized models, or simpler models. Also add regularization/early stopping.
Example: Switching from a single deep tree to a random forest stabilizes performance.

Q100: What is Rescaling of Data and how is it done?
A: Rescaling (feature scaling) puts features on a similar scale to help many models train better. Methods include standardization (z-score) and min-max scaling.
Example: Scale height (in cm) and weight (in kg) so SVM isn't dominated by larger-scale features.