# Section 2 — Machine Learning Theory

## Model Types & Use Cases

1. **What is supervised learning?**
   * ***Answer:* A learning approach where models are trained on labeled datasets to predict outcomes.**
2. **What is unsupervised learning?**
   * ***Answer:* A method where models find hidden patterns or groupings in unlabeled data.**
3. **What is reinforcement learning?**
   * ***Answer:* A learning paradigm where an agent interacts with an environment to maximize cumulative rewards.**
4. **Difference between classification and regression.**
   * ***Answer:* Classification predicts discrete categories, regression predicts continuous values.**
5. **When to use logistic regression vs decision trees?**
   * ***Answer:* Logistic regression works well for linearly separable problems; decision trees handle non-linear relationships and mixed data types.**
6. **What is clustering?**
   * ***Answer:* An unsupervised technique to group similar data points without predefined labels.**
7. **Difference between K-means and hierarchical clustering.**
   * ***Answer:* K-means partitions into k clusters, hierarchical clustering builds a tree of clusters.**
8. **What is dimensionality reduction?**
   * ***Answer:* Techniques (e.g., PCA) used to reduce the number of input variables while retaining essential information.**
9. **When to use PCA?**
   * ***Answer:* When features are highly correlated and you want to reduce noise and redundancy.**
10. **What is t-SNE used for?**
    * ***Answer:* A non-linear dimensionality reduction technique for visualizing high-dimensional data.**

## Feature Engineering

1. **What is one-hot encoding?**
   * ***Answer:* Converts categorical variables into binary indicator columns.**
2. **What is label encoding?**
   * ***Answer:* Assigns a unique integer to each category.**
3. **What is feature scaling?**
   * ***Answer:* Adjusting feature values to a similar range to improve model performance.**
4. **Difference between normalization and standardization.**
   * ***Answer:* Normalization rescales data to [0,1]; standardization centers data to mean=0 and std=1.**
5. **What is feature selection?**
   * ***Answer:* Choosing the most relevant features for model training to reduce overfitting and improve efficiency.**
6. **What is multicollinearity and how to handle it?**
   * ***Answer:* High correlation between features; can be addressed with PCA or dropping correlated features.**
7. **Explain polynomial feature generation.**
   * ***Answer:* Creating interaction and higher-order terms to capture non-linear relationships.**
8. **What is mean encoding?**
   * ***Answer:* Encoding categorical variables with the mean of the target variable for that category.**
9. **What is target leakage?**
   * ***Answer:* When data used for training includes information not available at prediction time.**
10. **How to handle missing data in ML?**
    * ***Answer:* Strategies include imputation (mean, median, mode), predictive models, or removal of missing rows.**

## Evaluation Metrics

1. **What is accuracy and when is it misleading?**
   * ***Answer:* Accuracy is the proportion of correct predictions; misleading with imbalanced datasets.**
2. **What is precision?**
   * ***Answer:* The proportion of positive predictions that are correct.**
3. **What is recall (sensitivity)?**
   * ***Answer:* The proportion of actual positives correctly identified.**
4. **What is F1-score?**
   * ***Answer:* Harmonic mean of precision and recall; useful when classes are imbalanced.**
5. **What is ROC curve?**
   * ***Answer:* A plot of true positive rate vs false positive rate at various thresholds.**
6. **What is AUC?**
   * ***Answer:* Area under the ROC curve; measures overall classification performance.**
7. **What is log loss?**
   * ***Answer:* Measures the uncertainty of predictions based on their probability estimates.**
8. **What is mean squared error (MSE)?**
   * ***Answer:* Average squared difference between predicted and actual values.**
9. **What is mean absolute error (MAE)?**
   * ***Answer:* Average absolute difference between predicted and actual values.**
10. **What is R-squared?**
    * ***Answer:* Proportion of variance explained by the model.**

## Overfitting & Regularization

1. **What is overfitting?**
   * ***Answer:* When a model learns noise instead of patterns, performing poorly on new data.**
2. **What is underfitting?**
   * ***Answer:* When a model is too simple to capture patterns in the data.**
3. **What is the bias-variance tradeoff?**
   * ***Answer:* Balance between underfitting (high bias) and overfitting (high variance).**
4. **What is L1 regularization?**
   * ***Answer:* Adds the absolute value of coefficients to the loss function; can drive some coefficients to zero.**
5. **What is L2 regularization?**
   * ***Answer:* Adds the squared value of coefficients to the loss; shrinks coefficients but doesn’t set them to zero.**
6. **What is elastic net?**
   * ***Answer:* Combines L1 and L2 regularization.**
7. **What is dropout?**
   * ***Answer:* Regularization in neural networks by randomly ignoring neurons during training.**
8. **What is early stopping?**
   * ***Answer:* Halting training when validation performance stops improving.**

## Ensemble Methods

1. **What is bagging?**
   * ***Answer:* Bootstrap aggregating; trains multiple models on random samples and averages predictions.**
2. **What is boosting?**
   * ***Answer:* Sequentially trains models, giving more weight to misclassified examples.**
3. **What is random forest?**
   * ***Answer:* Ensemble of decision trees using bagging and random feature selection.**
4. **What is gradient boosting?**
   * ***Answer:* Builds models sequentially to correct errors of previous models.**
5. **What is XGBoost?**
   * ***Answer:* An optimized gradient boosting library with regularization.**
6. **What is stacking?**
   * ***Answer:* Combining predictions of multiple models using a meta-model.**

## Deep Learning Basics

1. **What is a neural network?**
   * ***Answer:* A series of layers with interconnected nodes that learn representations of data.**
2. **What is backpropagation?**
   * ***Answer:* Algorithm for training neural networks by propagating error gradients backwards.**
3. **What is a convolutional neural network (CNN)?**
   * ***Answer:* Specialized neural network for processing grid-like data (e.g., images).**
4. **What is a recurrent neural network (RNN)?**
   * ***Answer:* Neural network designed for sequential data by maintaining hidden states.**
5. **What is long short-term memory (LSTM)?**
   * ***Answer:* A type of RNN that mitigates vanishing gradient problems with gating mechanisms.**
6. **What is a transformer model?**
   * ***Answer:* A deep learning architecture using self-attention, popular in NLP tasks.**

## Model Interpretability

1. **What is model interpretability?**
   * *Answer:* The degree to which a human can understand the internal mechanics and predictions of a machine learning model.
2. **What is SHAP?**
   * *Answer:* SHapley Additive exPlanations, a method to explain individual predictions based on game theory.
3. **What is LIME?**
   * *Answer:* Local Interpretable Model-agnostic Explanations, which explains predictions by approximating the model locally.
4. **What is partial dependence plot (PDP)?**
   * *Answer:* A visualization showing the effect of a feature on the predicted outcome, averaging over other features.
5. **What is a feature importance plot?**
   * *Answer:* A chart ranking features by their influence on model predictions.

## Hyperparameter Tuning

1. **What are hyperparameters?**
   * *Answer:* Parameters set before training that control the learning process (e.g., learning rate, tree depth).
2. **What is grid search?**
   * *Answer:* Exhaustive search over specified hyperparameter values.
3. **What is random search?**
   * *Answer:* Randomly sampling hyperparameters from given distributions.
4. **What is Bayesian optimization?**
   * *Answer:* Optimizing hyperparameters using a probabilistic model of the function mapping hyperparameters to performance.
5. **What is cross-validation in hyperparameter tuning?**
   * *Answer:* Splitting data into multiple folds to validate performance for each hyperparameter set.

## Model Deployment & Monitoring

1. **What is model deployment?**
   * *Answer:* Making a trained model available in a production environment for real-time or batch predictions.
2. **What is a REST API in ML deployment?**
   * *Answer:* An interface that allows applications to send requests to a deployed model and receive predictions.
3. **What is model versioning?**
   * *Answer:* Managing and tracking different versions of models to ensure reproducibility and rollback capability.
4. **What is model drift?**
   * *Answer:* Degradation in model performance due to changes in input data distribution or relationships.
5. **What is data drift?**
   * *Answer:* Change in the statistical properties of input data over time.
6. **What is concept drift?**
   * *Answer:* When the relationship between features and target changes over time.
7. **What is A/B testing for model deployment?**
   * *Answer:* Running two model versions in parallel to compare performance on live traffic.
8. **What is canary deployment?**
   * *Answer:* Rolling out a model to a small subset of users before full release.
9. **What is shadow deployment?**
   * *Answer:* Deploying a model alongside the current one to compare predictions without affecting live results.
10. **What is continuous training (CT) in MLOps?**
    * *Answer:* Automatically retraining and deploying models when new data becomes available.

## Advanced Topics — Reinforcement Learning

1. **What is Q-learning?**
   * *Answer:* A value-based RL algorithm where an agent learns a policy to maximize cumulative rewards using a Q-value function.
2. **What is the Bellman equation?**
   * *Answer:* A recursive formula that relates the value of a state to the values of successor states in RL.
3. **What is policy gradient?**
   * *Answer:* An RL method that directly optimizes the policy by adjusting parameters in the direction of performance improvement.
4. **What is an actor-critic method?**
   * *Answer:* Combines policy-based and value-based methods by maintaining both an actor (policy) and critic (value function).
5. **What is exploration vs exploitation tradeoff?**
   * *Answer:* The balance between trying new actions (exploration) and choosing the best-known action (exploitation).

## Advanced Topics — Generative Models

1. **What is a generative model?**
   * *Answer:* A model that learns the joint probability distribution of features and labels to generate new data points.
2. **What is a GAN (Generative Adversarial Network)?**
   * *Answer:* A framework with two networks (generator and discriminator) competing to generate realistic data.
3. **What is a variational autoencoder (VAE)?**
   * *Answer:* A generative model that learns a latent representation of input data using probabilistic encoders and decoders.
4. **What is diffusion modeling in ML?**
   * *Answer:* A generative approach that learns to reverse a gradual noising process to create new samples.
5. **What is the difference between conditional and unconditional generation?**
   * *Answer:* Conditional generation uses extra input (e.g., labels) to control generated outputs, unconditional does not.

## Advanced Topics — Ethical AI & Fairness

1. **What is bias in machine learning?**
   * *Answer:* Systematic error leading to unfair outcomes for certain groups.
2. **What is fairness in ML?**
   * *Answer:* Ensuring that a model's predictions are equitable across different subgroups.
3. **What is disparate impact?**
   * *Answer:* A metric that measures whether a decision disproportionately affects a protected group.
4. **What is explainable AI (XAI)?**
   * *Answer:* Techniques that make AI model decisions understandable to humans.
5. **What is adversarial attack in ML?**
   * *Answer:* Manipulating model inputs to cause incorrect predictions.
6. **What is adversarial training?**
   * *Answer:* Enhancing robustness by including adversarial examples in training.
7. **What is differential privacy?**
   * *Answer:* A technique that ensures the inclusion of a single data point does not significantly affect the output, protecting user privacy.
8. **What is federated learning?**
   * *Answer:* Training models collaboratively across devices without sharing raw data.

## Advanced Topics — Scalability & Optimization

1. **What is distributed training?**
   * *Answer:* Splitting model training across multiple machines or GPUs to handle large datasets.
2. **What is data parallelism vs model parallelism?**
   * *Answer:* Data parallelism splits data across workers, model parallelism splits the model itself.
3. **What is mixed precision training?**
   * *Answer:* Using both 16-bit and 32-bit floating-point operations to speed up training and save memory.
4. **What is model quantization?**
   * *Answer:* Reducing the precision of model weights to improve inference speed and reduce size.
5. **What is knowledge distillation?**
   * *Answer:* Training a smaller “student” model to replicate the behavior of a larger “teacher” model.
6. **What is transfer learning?**
   * *Answer:* Using a pretrained model as a starting point for a new task to save time and data.
7. **What is zero-shot learning?**
   * *Answer:* Predicting classes without having seen labeled examples for them during training.
8. **What is few-shot learning?**
   * *Answer:* Training a model to generalize to new classes given only a small number of examples.
9. **What is continual learning?**
   * *Answer:* Training models incrementally on new data while retaining knowledge from previous tasks.
10. **What is catastrophic forgetting?**
    * *Answer:* The tendency of neural networks to forget previously learned tasks when trained on new ones.
11. **What is meta-learning?**
    * *Answer:* “Learning to learn” — models improve their learning efficiency across tasks.
12. **What is neural architecture search (NAS)?**
    * *Answer:* Automatically finding the best neural network design for a given task.

# Section 10 — Data Technologies

1. **What is Apache Hadoop?**
   * *Answer:* An open-source framework for distributed storage and processing of large datasets using the MapReduce model.
2. **What is Apache Spark?**
   * *Answer:* A fast, in-memory distributed computing framework for big data analytics.
3. **What is Apache Flink used for?**
   * *Answer:* A stream-processing framework for high-throughput, low-latency data processing.
4. **What is Apache Kafka?**
   * *Answer:* A distributed messaging and streaming platform for real-time data pipelines.
5. **What is Elasticsearch?**
   * *Answer:* A distributed search and analytics engine used for log analytics, full-text search, and more.
6. **What is MongoDB?**
   * *Answer:* A NoSQL document-oriented database for storing unstructured data.
7. **What is Cassandra?**
   * *Answer:* A distributed NoSQL database designed for scalability and high availability.
8. **What is Snowflake?**
   * *Answer:* A cloud-based data warehouse platform with separation of storage and compute.
9. **What is Google BigQuery?**
   * *Answer:* A serverless, cloud-based enterprise data warehouse for fast SQL queries.
10. **What is Amazon Redshift?**
    * *Answer:* A cloud data warehouse service from AWS optimized for analytics.
11. **What is Databricks?**
    * *Answer:* A cloud platform for big data analytics and AI, built around Apache Spark.
12. **What is Airflow used for?**
    * *Answer:* Workflow orchestration and scheduling for data pipelines.
13. **What is dbt (data build tool)?**
    * *Answer:* A tool for transforming data inside warehouses using SQL.
14. **What is Tableau?**
    * *Answer:* A business intelligence platform for interactive data visualization.
15. **What is Power BI?**
    * *Answer:* Microsoft’s data visualization and BI platform.
16. **What is Looker?**
    * *Answer:* A modern BI and analytics platform with a modeling layer for data.
17. **What is OLAP?**
    * *Answer:* Online Analytical Processing—used for complex analytical queries on multidimensional data.
18. **What is OLTP?**
    * *Answer:* Online Transaction Processing—used for fast, transactional workloads.
19. **What is a data catalog?**
    * *Answer:* A centralized inventory of data assets for discovery and governance.
20. **What is a data mesh?**
    * *Answer:* A decentralized data architecture focusing on domain-oriented ownership and self-serve infrastructure.
21. **What is a data fabric?**
    * *Answer:* An integrated layer of data and connecting processes providing consistent data services across environments.
22. **What is Presto/Trino?**
    * *Answer:* Distributed SQL query engines for querying data across multiple sources.
23. **What is Apache Superset?**
    * *Answer:* An open-source data exploration and visualization platform.
24. **What is Matillion?**
    * *Answer:* A cloud-native ETL/ELT tool for data transformation.
25. **What is Fivetran?**
    * *Answer:* A managed ELT service for automated data integration.
26. **What is Talend?**
    * *Answer:* An open-source data integration and transformation platform.
27. **What is Informatica?**
    * *Answer:* A data integration and management platform.
28. **What is Alteryx?**
    * *Answer:* A platform for data preparation, blending, and advanced analytics.
29. **What is RapidMiner?**
    * *Answer:* A data science platform for building predictive models.
30. **What is KNIME?**
    * *Answer:* An open-source platform for data analytics, reporting, and integration.
31. **What is a vector database?**
    * *Answer:* A specialized database optimized for storing and querying vector embeddings, used in AI search and recommendation systems.
32. **What is Pinecone?**
    * *Answer:* A managed vector database service for building semantic search and recommendation applications.
33. **What is Weaviate?**
    * *Answer:* An open-source vector search engine that integrates machine learning models for semantic search.
34. **What is Milvus?**
    * *Answer:* An open-source vector database designed for scalable similarity search.
35. **What is DuckDB?**
    * *Answer:* An in-process analytical database optimized for OLAP-style queries on local data.
36. **What is ClickHouse?**
    * *Answer:* A columnar database management system optimized for real-time analytics.
37. **What is TimescaleDB?**
    * *Answer:* A time-series database built on PostgreSQL, optimized for storing and querying time-based data.
38. **What is InfluxDB?**
    * *Answer:* An open-source time-series database for metrics and events.
39. **What is Apache Druid?**
    * *Answer:* A real-time analytics database for high-performance OLAP queries.
40. **What is Neo4j?**
    * *Answer:* A graph database used for storing and querying data with complex relationships.
41. **What is JanusGraph?**
    * *Answer:* An open-source distributed graph database optimized for large-scale graph processing.
42. **What is Redis?**
    * *Answer:* An in-memory data structure store used as a database, cache, and message broker.
43. **What is Hazelcast?**
    * *Answer:* An in-memory data grid for fast data access and distributed computing.
44. **What is Apache Ignite?**
    * *Answer:* An in-memory computing platform for high-performance processing and distributed storage.
45. **What is DataRobot?**
    * *Answer:* An automated machine learning (AutoML) platform for building and deploying predictive models.
46. **What is H2O.ai?**
    * *Answer:* An open-source machine learning and AI platform with AutoML capabilities.
47. **What is Vertex AI?**
    * *Answer:* Google Cloud’s managed ML platform for building, training, and deploying AI models.
48. **What is SageMaker?**
    * *Answer:* AWS’s managed service for building, training, and deploying machine learning models.
49. **What is Azure Machine Learning?**
    * *Answer:* Microsoft Azure’s platform for end-to-end machine learning lifecycle management.
50. **What is Feast?**
    * *Answer:* An open-source feature store for managing and serving machine learning features.
51. **What is the difference between OLTP and OLAP systems?**
    * *Answer:* OLTP is optimized for transaction-oriented tasks with frequent reads/writes, while OLAP is optimized for complex analytical queries on aggregated historical data.
52. **What is ACID compliance in databases?**
    * *Answer:* A set of properties—Atomicity, Consistency, Isolation, Durability—ensuring reliable transaction processing.
53. **What is BASE in NoSQL databases?**
    * *Answer:* An alternative to ACID, meaning Basically Available, Soft state, Eventually consistent, common in distributed systems.
54. **What is eventual consistency?**
    * *Answer:* A consistency model in distributed databases where updates will propagate to all nodes over time.
55. **What is sharding?**
    * *Answer:* The practice of splitting a large dataset into smaller, faster, more easily managed parts called shards.
56. **What is data partitioning?**
    * *Answer:* Dividing a database into distinct independent parts to improve performance, manageability, or availability.
57. **What is replication in databases?**
    * *Answer:* The process of copying and maintaining database objects in multiple locations to improve availability and reliability.
58. **What is a CAP theorem?**
    * *Answer:* States that in distributed systems, you can only guarantee two of the following three: Consistency, Availability, and Partition tolerance.
59. **What is a data warehouse schema?**
    * *Answer:* The logical structure of data storage, e.g., star schema, snowflake schema, and galaxy schema.
60. **What is a star schema?**
    * *Answer:* A data warehouse schema with a central fact table connected to dimension tables.
61. **What is a snowflake schema?**
    * *Answer:* A normalized version of the star schema where dimension tables are split into related tables.
62. **What is a fact table in data warehousing?**
    * *Answer:* A table containing quantitative data for analysis, linked to dimension tables.
63. **What are dimension tables?**
    * *Answer:* Tables in a data warehouse containing descriptive attributes related to fact data.
64. **What is data normalization?**
    * *Answer:* The process of structuring a relational database to reduce redundancy and improve data integrity.
65. **What is data denormalization?**
    * *Answer:* The process of adding redundancy to a database to improve read performance.
66. **What is an ETL process?**
    * *Answer:* Extract, Transform, Load—moving and transforming data from sources into a target system like a data warehouse.
67. **What is an ELT process?**
    * *Answer:* Extract, Load, Transform—loading raw data into the target system first, then transforming it.
68. **What is batch processing?**
    * *Answer:* Processing data in large groups at scheduled intervals.
69. **What is stream processing?**
    * *Answer:* Processing data in real time as it is ingested.
70. **What is data lineage?**
    * *Answer:* The record of the data’s origins, movement, and transformations across systems.

# Section 13 — ICT (Information and Communication Technology) in Data Science

1. **What is ICT in the context of data science?**
   * *Answer:* The integration of computing, communication, and data processing technologies to enable data-driven decision-making.
2. **How does ICT support data science workflows?**
   * *Answer:* By providing infrastructure, networking, communication channels, and tools for collaboration and data sharing.
3. **What are examples of ICT infrastructure in data projects?**
   * *Answer:* Data centers, cloud platforms, high-speed networks, and collaboration tools like Slack or Microsoft Teams.
4. **What is the role of networking in big data processing?**
   * *Answer:* High-bandwidth, low-latency networks facilitate rapid data transfer between storage, processing, and visualization systems.
5. **How does ICT enable remote data science teams?**
   * *Answer:* Through cloud services, secure VPNs, shared code repositories, and real-time communication platforms.
6. **What are key ICT security considerations in data projects?**
   * *Answer:* Access control, encryption, secure communication protocols, and endpoint security.
7. **How do collaboration platforms impact data science productivity?**
   * *Answer:* They improve coordination, reduce delays, and centralize discussions and documentation.
8. **What is unified communications in ICT?**
   * *Answer:* The integration of multiple communication methods—voice, video, messaging—into a single platform.
9. **How does ICT integrate with MLOps pipelines?**
   * *Answer:* By supporting CI/CD tools, monitoring systems, and secure data transfer between environments.
10. **What are ICT standards relevant to data science teams?**
    * *Answer:* ISO/IEC 27001 for security, ITIL for service management, and IEEE standards for interoperability.
11. **How does ICT support real-time analytics?**
    * *Answer:* By enabling low-latency data streaming, processing, and visualization across distributed systems.
12. **What is the role of APIs in ICT for data science?**
    * *Answer:* APIs facilitate interoperability between systems, allowing seamless integration of tools and services.
13. **How does ICT enable data democratization?**
    * *Answer:* Through shared platforms, self-service analytics tools, and secure role-based access to data.
14. **What ICT solutions help manage large-scale datasets?**
    * *Answer:* Distributed file systems, cloud storage, and high-performance computing clusters.
15. **How do ICT advancements influence AI deployment?**
    * *Answer:* Faster networks, edge computing, and scalable infrastructure improve AI model performance and accessibility.
16. **What is the importance of ICT policy in organizations?**
    * *Answer:* Policies govern technology use, security compliance, and resource allocation for consistent and safe operations.
17. **How do ICT tools assist in cross-border data projects?**
    * *Answer:* They enable secure communication, manage time zone differences, and comply with regional data laws.
18. **What is the role of ICT in disaster recovery for data systems?**
    * *Answer:* ICT ensures backup systems, failover capabilities, and business continuity plans are in place.
19. **How does ICT impact the scalability of data science solutions?**
    * *Answer:* Robust ICT infrastructure supports the growth of processing capacity, user access, and data storage.
20. **What trends in ICT are shaping the future of data science?**
    * *Answer:* 5G networks, quantum computing, AI-driven automation, and expanded use of edge computing.
21. **What is ICT architecture in data science environments?**
    * *Answer:* The structured design of hardware, software, networking, and data systems that support analytics and AI workflows.
22. **What is interoperability in ICT systems?**
    * *Answer:* The ability of different ICT systems, applications, and components to communicate and work together seamlessly.
23. **What is the role of virtualization in ICT infrastructure?**
    * *Answer:* Virtualization enables multiple virtual environments to run on a single physical system, improving flexibility and resource utilization.
24. **What is containerization in ICT for data science?**
    * *Answer:* Packaging applications and their dependencies into containers for portability, scalability, and consistency.
25. **How does ICT support hybrid cloud strategies?**
    * *Answer:* By enabling secure integration and management of workloads across on-premise, private, and public cloud environments.
26. **What are ICT redundancy strategies?**
    * *Answer:* Methods like load balancing, failover, and replication to ensure continuous service availability.
27. **What is the role of ICT in edge computing for AI?**
    * *Answer:* Deploying models close to data sources to reduce latency and bandwidth usage.
28. **What is ICT orchestration?**
    * *Answer:* Automated coordination and management of interconnected systems, services, and processes.
29. **What are ICT service-level agreements (SLAs)?**
    * *Answer:* Formal agreements defining expected service performance, availability, and responsibilities.
30. **What is ICT change management?**
    * *Answer:* Processes to control and document changes in ICT systems to minimize disruption and risk.
31. **How does ICT support multi-tenancy in data platforms?**
    * *Answer:* By isolating workloads and ensuring secure resource sharing for multiple users or organizations.
32. **What is the difference between synchronous and asynchronous communication in ICT?**
    * *Answer:* Synchronous occurs in real time (e.g., video calls); asynchronous allows delayed responses (e.g., email, message boards).
33. **What is ICT capacity planning?**
    * *Answer:* Predicting and preparing for future infrastructure requirements to handle growth in users and data volume.
34. **What is network segmentation in ICT security?**
    * *Answer:* Dividing a network into smaller, isolated segments to limit unauthorized access and improve security.
35. **What is ICT’s role in high-performance computing (HPC) for data science?**
    * *Answer:* Providing the infrastructure to execute large-scale, compute-intensive workloads efficiently.
36. **What is ICT convergence?**
    * *Answer:* The integration of computing, networking, and content delivery into a unified system.
37. **What is the importance of ICT compliance monitoring?**
    * *Answer:* Ensures systems adhere to legal, regulatory, and internal policy requirements.
38. **What are ICT green computing practices?**
    * *Answer:* Techniques to reduce environmental impact, such as energy-efficient hardware and optimized cooling systems.
39. **What is ICT’s role in secure API management?**
    * *Answer:* Governing, monitoring, and securing APIs to ensure safe data exchange between systems.
40. **What is ICT performance monitoring?**
    * *Answer:* Tracking system metrics like uptime, latency, and throughput to maintain optimal performance.
41. **What are the main types of system architectures in ICT for data science?**
    * *Answer:* Common types include monolithic, layered (n-tier), microservices, event-driven, and serverless architectures.
42. **What is a layered architecture?**
    * *Answer:* An architecture where the system is divided into layers such as presentation, application, and data, each with defined responsibilities.
43. **What is a microservices architecture?**
    * *Answer:* An approach where the application is built as a collection of loosely coupled services that can be developed, deployed, and scaled independently.
44. **What is an event-driven architecture?**
    * *Answer:* A design where system components communicate through events, improving responsiveness and scalability.
45. **What is a serverless architecture?**
    * *Answer:* A model where developers focus on writing code without managing servers, with execution handled by cloud providers.
46. **What is a data lake architecture?**
    * *Answer:* A centralized repository that stores raw data in its native format until it is needed for analysis.
47. **What is a data warehouse architecture?**
    * *Answer:* A structured system optimized for reporting and analysis, often organized into staging, integration, and presentation layers.
48. **What is a lakehouse architecture?**
    * *Answer:* A hybrid architecture that combines the flexibility of data lakes with the management features of data warehouses.
49. **What is a hub-and-spoke architecture in data platforms?**
    * *Answer:* A central data hub integrates and governs data, while spokes represent domain-specific data marts.
50. **What is edge architecture in AI systems?**
    * *Answer:* A system design where data processing occurs close to the data source, reducing latency and bandwidth usage.
51. **What is cloud computing and its main service models?**
    * *Answer:* Delivery of computing services over the internet; main models are IaaS (Infrastructure as a Service), PaaS (Platform as a Service), and SaaS (Software as a Service).
52. **What is the difference between public, private, and hybrid clouds?**
    * *Answer:* Public clouds are shared infrastructure managed by providers, private clouds are dedicated to one organization, and hybrid clouds combine both.
53. **What is IoT in the context of ICT?**
    * *Answer:* The Internet of Things connects devices and sensors to collect, exchange, and process data.
54. **How are drones used in ICT-enabled systems?**
    * *Answer:* For aerial data collection, inspections, delivery services, and environmental monitoring.
55. **What types of sensors are common in IoT?**
    * *Answer:* Temperature, humidity, motion, GPS, proximity, and biometric sensors.
56. **What is M2M communication?**
    * *Answer:* Machine-to-machine communication enables devices to exchange data without human intervention.
57. **What is 5G technology and its relevance to ICT?**
    * *Answer:* The fifth generation of mobile networks, offering high speed, low latency, and supporting massive IoT connectivity.
58. **What are common communication protocols in ICT?**
    * *Answer:* HTTP/HTTPS, MQTT, CoAP, FTP, TCP/IP, and WebSocket.
59. **What is robotics in the ICT domain?**
    * *Answer:* The design, construction, and operation of programmable machines to automate tasks.
60. **What are industrial robots and collaborative robots (cobots)?**
    * *Answer:* Industrial robots operate in controlled environments; cobots are designed to work alongside humans safely.
61. **What is an embedded system?**
    * *Answer:* A computer system with a dedicated function within a larger system, often with real-time constraints.
62. **What is firmware in ICT systems?**
    * *Answer:* Software embedded in hardware to control device functions.
63. **What is edge AI?**
    * *Answer:* Running AI algorithms directly on devices at the network edge for faster, offline processing.
64. **What are real-time operating systems (RTOS)?**
    * *Answer:* Operating systems designed for applications requiring precise timing and reliability.
65. **What is platform as a service (PaaS) in ICT?**
    * *Answer:* A cloud service model providing infrastructure and tools to develop, test, and deploy applications.
66. **What is SaaS and give examples relevant to data science?**
    * *Answer:* Software as a Service delivers applications over the internet; examples: Google BigQuery, Tableau Online.
67. **What are distributed algorithms?**
    * *Answer:* Algorithms designed to run on multiple interconnected computers, coordinating actions and data exchange.
68. **What is parallel computing?**
    * *Answer:* Simultaneous execution of computations across multiple processors to increase performance.
69. **What is quantum computing and its ICT relevance?**
    * *Answer:* Computing using quantum bits to solve certain problems faster than classical computers.
70. **What is augmented reality (AR) and virtual reality (VR) in ICT?**
    * *Answer:* AR overlays digital content on the real world; VR creates immersive simulated environments
71. **What is cybersecurity in ICT?**
    * *Answer:* The practice of protecting systems, networks, and data from digital attacks, damage, or unauthorized access.
72. **What is network infrastructure?**
    * *Answer:* The hardware and software resources enabling network connectivity, communication, and operations.
73. **What are the main types of firewalls?**
    * *Answer:* Packet-filtering, stateful inspection, proxy, and next-generation firewalls.
74. **What is a VPN and its purpose?**
    * *Answer:* A Virtual Private Network encrypts internet connections, providing secure remote access.
75. **What are common types of malware?**
    * *Answer:* Viruses, worms, trojans, ransomware, spyware, and adware.
76. **What is the difference between a virus and a worm?**
    * *Answer:* Viruses require a host file to spread; worms are self-replicating and spread without user action.
77. **What is phishing?**
    * *Answer:* A cyberattack method where attackers deceive individuals into revealing sensitive information.
78. **What is a denial-of-service (DoS) attack?**
    * *Answer:* An attack that overwhelms a system or network to make it unavailable to users.
79. **What is intrusion detection and prevention (IDS/IPS)?**
    * *Answer:* Systems that monitor network traffic for suspicious activity and take action to block threats.
80. **What is encryption and why is it important?**
    * *Answer:* Encoding data to protect its confidentiality and integrity during storage and transmission.
81. **What is public key infrastructure (PKI)?**
    * *Answer:* A system for creating, managing, and validating digital certificates for secure communications.
82. **What is a man-in-the-middle (MITM) attack?**
    * *Answer:* When an attacker intercepts communication between two parties without their knowledge.
83. **What is endpoint security?**
    * *Answer:* Protection of devices like laptops, phones, and IoT gadgets from cyber threats.
84. **What is network segmentation and why is it important?**
    * *Answer:* Dividing a network into subnets to improve performance and security.
85. **What is zero trust security?**
    * *Answer:* A security approach that assumes no implicit trust and requires continuous verification.
86. **What is social engineering in cybersecurity?**
    * *Answer:* Manipulating people into performing actions or revealing confidential information.
87. **What are honeypots in network security?**
    * *Answer:* Decoy systems designed to lure and analyze attackers’ behavior.
88. **What is network latency and why does it matter?**
    * *Answer:* The delay between data transmission and reception; impacts performance.
89. **What are intrusion prevention best practices?**
    * *Answer:* Regular patching, strong authentication, network monitoring, and user training.
90. **What is ethical hacking?**
    * *Answer:* Authorized testing of systems to find and fix security vulnerabilities before malicious actors exploit them.

# Section 16 — Communication Technologies and Systems

1. **What are the main components of a communication system?**
   * *Answer:* Transmitter, transmission medium, receiver, and protocols governing the exchange.
2. **What is the difference between analog and digital communication?**
   * *Answer:* Analog transmits continuous signals; digital transmits discrete binary signals.
3. **What are the OSI model layers?**
   * *Answer:* Physical, Data Link, Network, Transport, Session, Presentation, and Application layers.
4. **What is TCP/IP and why is it important?**
   * *Answer:* A suite of communication protocols used to interconnect network devices on the internet.
5. **What is 5G technology and its relevance to communication systems?**
   * *Answer:* Fifth-generation mobile networks offering higher speeds, low latency, and support for massive IoT connectivity.
6. **What is the difference between synchronous and asynchronous communication?**
   * *Answer:* Synchronous requires simultaneous presence of sender and receiver; asynchronous allows delayed responses.
7. **What is VoIP and how does it work?**
   * *Answer:* Voice over Internet Protocol transmits voice data over IP networks instead of traditional telephony.
8. **What is MIMO in wireless communication?**
   * *Answer:* Multiple Input Multiple Output uses multiple antennas for improved performance and capacity.
9. **What is satellite communication?**
   * *Answer:* Data transmission via satellites for global coverage, often used in remote areas.
10. **What are common communication protocols in IoT systems?**
    * *Answer:* MQTT, CoAP, Zigbee, LoRaWAN, and Bluetooth Low Energy.
11. **What is network latency and how can it be reduced?**
    * *Answer:* Delay between sending and receiving data; reduced with faster hardware, optimized routing, and edge computing.
12. **What is the difference between unicast, multicast, and broadcast transmission?**
    * *Answer:* Unicast sends data to one recipient, multicast to a group, and broadcast to all nodes in a network.
13. **What is fiber optic communication and its advantages?**
    * *Answer:* Data transmission via light signals in optical fibers; offers high speed, long distance, and resistance to interference.
14. **What is network topology?**
    * *Answer:* The arrangement of network elements—common types include star, mesh, bus, and ring.
15. **What are APIs in communication systems?**
    * *Answer:* Application Programming Interfaces enabling integration and data exchange between systems.
16. **What is edge computing in communication networks?**
    * *Answer:* Processing data near the source to reduce latency and bandwidth use.
17. **What is the role of DNS in internet communication?**
    * *Answer:* Translates human-readable domain names into IP addresses.
18. **What is network security in communication systems?**
    * *Answer:* Measures to protect data and resources from unauthorized access or attacks.
19. **What is Unified Communications (UC)?**
    * *Answer:* The integration of multiple communication methods (voice, video, messaging) into a single platform.
20. **What is Quality of Service (QoS) in networking?**
    * *Answer:* Techniques to prioritize certain types of network traffic for performance assurance.

# Section 17 — Information Technologies and Systems

1. **What is Information Technology (IT) in the context of ICT?**
   * *Answer:* IT refers to the use of computers, networks, storage, and other physical devices to process, store, secure, and exchange electronic data.
2. **What are the core components of an information system?**
   * *Answer:* Hardware, software, data, people, and processes.
3. **What is the difference between IT and ICT?**
   * *Answer:* IT focuses on computing and data management; ICT encompasses IT plus communication technologies.
4. **What is an ERP system?**
   * *Answer:* Enterprise Resource Planning integrates core business processes like finance, HR, and supply chain into a unified system.
5. **What is a CRM system?**
   * *Answer:* Customer Relationship Management systems manage interactions with current and potential customers.
6. **What is the role of IT governance?**
   * *Answer:* Ensures IT systems align with business goals, manage risks, and deliver value.
7. **What is cloud-based information system deployment?**
   * *Answer:* Hosting information systems on cloud infrastructure for scalability and accessibility.
8. **What is a data management system?**
   * *Answer:* Software for creating, retrieving, updating, and managing data, such as databases and data warehouses.
9. **What is a knowledge management system (KMS)?**
   * *Answer:* A platform for capturing, storing, sharing, and managing organizational knowledge.
10. **What is business intelligence (BI) in IT systems?**
    * *Answer:* Technologies and tools for analyzing business data to support decision-making.
11. **What is IT service management (ITSM)?**
    * *Answer:* A set of processes for delivering and managing quality IT services.
12. **What is an MIS (Management Information System)?**
    * *Answer:* An IT system providing information for managing an organization effectively.
13. **What is a decision support system (DSS)?**
    * *Answer:* IT-based tools that help in making informed business or operational decisions.
14. **What is virtualization in IT systems?**
    * *Answer:* Creating virtual versions of computing resources, improving flexibility and utilization.
15. **What is a disaster recovery plan (DRP) in IT?**
    * *Answer:* A documented process to restore IT systems and data after an outage or disaster.
16. **What is cybersecurity's role in IT systems?**
    * *Answer:* Protecting IT assets and data from breaches, attacks, or unauthorized access.
17. **What is the role of APIs in IT systems integration?**
    * *Answer:* APIs allow different software systems to exchange data and functionality.
18. **What is IT compliance and why is it important?**
    * *Answer:* Adhering to legal, regulatory, and industry standards for IT operations.
19. **What is the role of AI in IT systems?**
    * *Answer:* Automating tasks, enhancing analytics, and enabling intelligent decision-making.
20. **What trends are shaping the future of IT systems?**
    * *Answer:* Cloud-native architectures, edge computing, AI integration, automation, and zero trust security models.

# Section 18 — Network Technologies and Systems

1. **What are network technologies in the context of ICT?**
   * *Answer:* The hardware, software, and protocols that enable data transmission, connectivity, and communication between systems.
2. **What are the main types of computer networks?**
   * *Answer:* LAN (Local Area Network), WAN (Wide Area Network), MAN (Metropolitan Area Network), and PAN (Personal Area Network).
3. **What is the difference between circuit-switched and packet-switched networks?**
   * *Answer:* Circuit-switched networks establish a dedicated communication path; packet-switched networks send data in packets over shared paths.
4. **What is Ethernet?**
   * *Answer:* A widely used wired networking technology for local area networks.
5. **What is Wi-Fi and how does it work?**
   * *Answer:* A wireless networking technology using radio waves to provide high-speed internet and network connections.
6. **What is network topology?**
   * *Answer:* The physical or logical arrangement of network devices and connections (e.g., star, bus, mesh, ring).
7. **What is the difference between IPv4 and IPv6?**
   * *Answer:* IPv4 uses 32-bit addresses; IPv6 uses 128-bit addresses for a vastly larger address space.
8. **What are network protocols?**
   * *Answer:* Rules and conventions for communication between network devices (e.g., TCP/IP, HTTP, FTP, SNMP).
9. **What is a VLAN and why is it used?**
   * *Answer:* Virtual LAN segments a physical network into logical parts for better management and security.
10. **What is network segmentation?**
    * *Answer:* Dividing a network into multiple segments to improve performance and security.
11. **What is SDN (Software-Defined Networking)?**
    * *Answer:* An approach that separates the control plane from the data plane for more flexible network management.
12. **What is NFV (Network Functions Virtualization)?**
    * *Answer:* Replacing dedicated network hardware with virtualized functions running on standard servers.
13. **What is a network switch?**
    * *Answer:* A device that connects devices within a LAN and forwards data based on MAC addresses.
14. **What is a router and its function?**
    * *Answer:* A device that directs data between networks based on IP addresses.
15. **What is a firewall in networking?**
    * *Answer:* A security system that monitors and controls incoming and outgoing network traffic.
16. **What is network redundancy and why is it important?**
    * *Answer:* Using backup network paths and devices to ensure reliability and uptime.
17. **What is a load balancer?**
    * *Answer:* A device or software that distributes network traffic across multiple servers for efficiency and reliability.
18. **What is a content delivery network (CDN)?**
    * *Answer:* A distributed network of servers that deliver web content based on user location.
19. **What is network monitoring?**
    * *Answer:* Continuous observation of network traffic and performance to detect and resolve issues.
20. **What trends are shaping the future of network technologies?**
    * *Answer:* 5G/6G networks, edge networking, intent-based networking, and AI-driven network management.