
All existing questions collection of IS 466 – Decision Support System

until Fall-2024

Chapter 1, 2, 4 and 7.

Chapter 1:

A decision support system is needed to solve which type of problems?

- A) Structured
- B) Semi-structured
- C) Unstructured
- D) Both Semi-structured and unstructured

 **Answer: D) Both Semi-structured and unstructured**

In the decision-making process defined by Quain (2018), which of the following should be the first step to make?

- A) Collect the information
- B) Select the best alternative
- C) Understand the decision you have to make
- D) Evaluate the impact of your decision

 **Answer: C) Understand the decision you have to make**

Which of the following is the oldest tool/concepts used as decision support systems?

- A) Business Intelligence
- B) Analytics
- C) Expert Systems
- D) Executive Information Systems

 **Answer: C) Expert Systems**

Which of the following is an activity that typically requires managerial control?

- A) Monitoring accounts receivables
- B) Scheduling production
- C) Scheduling project
- D) Building a new plant

✓ **Answer: C) Scheduling project**

OLAP is a type of processing that is typically used with which of the following?

- A) Data Warehouse
- B) Expert System
- C) ERP
- D) MIS

✓ **Answer: A) Data Warehouse**

What type of analytics answers questions like "what will happen?"

- A) Business intelligence
- B) Descriptive
- C) Predictive
- D) Prescriptive

✓ **Answer: C) Predictive**

Knowing what is happening in the organization and understanding some underlying trends and causes of such occurrences is achieved by:

- A) Business intelligence
- B) Descriptive analytics
- C) Predictive analytics
- D) Prescriptive analytics

✓ **Answer: B) Descriptive analytics**

Which of the following is NOT a main science or tool used in Analytics?

- A) Cognitive computing
- B) Data warehouse
- C) Statistics
- D) Management science

✓ **Answer: A) Cognitive computing**

Which of the following is NOT one of the known AI applications?

- A) Autonomous vehicles
- B) Recommender systems
- C) Sensors
- D) Fraud detection

✓ **Answer: C) Sensors**

Alexa and Siri are examples of what type of AI?

- A) All options are true
- B) Knowledge-based systems
- C) Personalized services
- D) Voice recognition

✓ **Answer: A) All options are true**

Cybercrime fighting is considered under what flavor of AI?

- A) None of the options is true
- B) Assisted AI
- C) Autonomous AI
- D) Augmented Intelligence

✓ **Answer: D) Augmented Intelligence**

Machines that will have decision-making power are under what flavor of AI?

- A) None of the options is true
- B) Assisted AI
- C) Autonomous AI
- D) Augmented Intelligence

✓ **Answer: C) Autonomous AI**

Enterprises have voiced that 70-80% of analytics initiatives will not meet their objectives, and hence, the best solution is to combine both analytics and AI to improve decision-making and get better results.

- A) True
- B) False

✓ **Answer: A) True**

In the decision-making process proposed by Simon (1977), which of the following should be the second phase?

- A) Implementation
- B) Intelligence
- C) Design
- D) Choice

✓ **Answer: C) Design**

A system that relies mostly on heuristics (what-if-then):

- A) Enterprise Resource Planning
- B) Expert System
- C) Executive Information Systems
- D) Data Warehousing

✓ **Answer: B) Expert System**

A system that relies mostly on historical and comprehensive data:

- A) Enterprise Resource Planning
- B) Expert System
- C) Executive Information Systems
- D) Data Warehousing

✓ **Answer: D) Data Warehousing**

A system that relies mostly on providing dashboards and easy-to-understand graphics:

- A) Enterprise Resource Planning
- B) Expert System
- C) Executive Information Systems
- D) Data Warehousing

✓ **Answer: C) Executive Information Systems**

A system that creates a single source of truth for data items on daily functional activities:

- A) Enterprise Resource Planning
- B) Expert System
- C) Executive Information Systems
- D) Data Warehousing

✓ **Answer: A) Enterprise Resource Planning**

One of the most important systems that make business intelligence possible is:

- A) Enterprise Resource Planning
- B) Expert System
- C) Executive Information Systems
- D) Data Warehousing

✓ **Answer: D) Data Warehousing**

What is a data mart?

- A) A web-based database where clients can see their transaction history
- B) A system where companies are able to buy data about customers
- C) A smaller subset of a data warehouse dedicated for a special business function
- D) A much larger system than a data warehouse that includes external data

✓ **Answer: C) A smaller subset of a data warehouse dedicated for a special business function**

The letters in the ETL process stand for:

- A) Extract, Transform, Load
- B) Enhance, Toggle, Load
- C) Extend, Track, Load
- D) Extract, Terminate, Load

✓ **Answer: A) Extract, Transform, Load**

The purpose of descriptive analytics is to:

- A) Give recommendations on what actions to take
- B) Determine the expected future outcome of events
- C) Describe the current situation
- D) Take automated actions driven by AI

✓ **Answer: C) Describe the current situation**

The purpose of predictive analytics is to:

- A) Give recommendations on what actions to take
- B) Determine the expected future outcome of events
- C) Describe the current situation
- D) Take automated actions driven by AI

✓ **Answer: B) Determine the expected future outcome of events**

What type of analytics answers questions like "what will happen in the future?"

- A) Business intelligence
- B) Descriptive
- C) Predictive
- D) Prescriptive

✓ **Answer: C) Predictive**

What is a data mart?

- A) A much larger system than a data warehouse that includes external data
- B) A web-based database where clients can see their transaction history
- C) A smaller subset of a data warehouse dedicated for a special business function
- D) A system where companies are able to buy data about customers

✓ Answer: C) A smaller subset of a data warehouse dedicated for a special business function

The letters in the ETL process stand for:

- A) Extract, Terminate, Load
- B) Extract, Transform, Load
- C) Enhance, Toggle, Load
- D) Extend, Track, Load

✓ Answer: B) Extract, Transform, Load

What does a Hadoop Distributed File System do?

- A) Uses Super Computers to process Big Data
- B) Uses servers/CPUs over the network to process Big Data
- C) Uses different machines over the network to store Big Data
- D) All of the above

✓ Answer: C) Uses different machines over the network to store Big Data

Machines that have decision-making power are under what flavor of AI?

- A) Augmented Intelligence
- B) Assisted AI
- C) Autonomous AI
- D) None of the options is true

✓ Answer: C) Autonomous AI

Why is there a big need for decision support systems?

- A) All following options are true
- B) Decisions with big impact and potential high risks
- C) Sudden unanticipated issues
- D) Issues that require collaboration across different departments

✓ Answer: A) All following options are true

Which of the following is a step among the decision-making process?

- A) All following options are true
- B) Understand the exact problem you are dealing with
- C) Preview previous records and related files
- D) Analyze the potential negatives and positives of your potential decision

✓ **Answer: A) All following options are true**

An example of a factor that may result in the need for the company's upper management to make an important strategic decision:

- A) All following options are true
- B) Company's CEO has decided to leave to work for another company
- C) Sales have been declining on a monthly basis for the last 6 months
- D) Company's main competitor has released a new line of trendy products

✓ **Answer: A) All following options are true**

An example of a global factor that may impact a company's production cost:

- A) All following options are true
- B) The decrease in the number of local male customers
- C) The rise of unemployment rate in a neighboring country
- D) The increase in overseas shipping costs

✓ **Answer: D) The increase in overseas shipping costs**

Which of the following is a type of decision that can be easily and fully dealt with using typical information systems?

- A) All following options are true
- B) Unstructured decisions
- C) Semi-structured decisions
- D) Structured decisions

✓ **Answer: D) Structured decisions**

Which of the following is NOT a key characteristic or capability of a DSS?

- A) Enables easy data access
- B) May be developed as a web-based tool
- C) Enables daily functional transaction processing
- D) Enables dealing with unstructured problems

✓ **Answer: C) Enables daily functional transaction processing**

Which of the following is NOT a component of a DSS?

- A) Database management system
- B) User interface
- C) Knowledge-based subsystems
- D) Enterprise Resource Planning system

✓ **Answer: D) Enterprise Resource Planning system**

Which is NOT a component of a model management subsystem?

- A) External models subsystem
- B) Model execution, integration, and command processor
- C) Model directory
- D) Model base management system

✓ **Answer: A) External models subsystem**

Which of the following is a problem that ERP systems solved?

- A) All of the following options are true
- B) Inability to run queries on the database
- C) Inability to create daily reports
- D) Inability to access data from different departments

✓ **Answer: D) Inability to access data from different departments**

Which of the following describes advantages of software as a service (SaaS)?

- A) All of the following options are true
- B) Relieves a company from the need to own a software
- C) Is very suitable for small and medium enterprises (SMEs)
- D) Is accessed through the cloud

✓ **Answer: A) All of the following options are true**

In the decision-making process defined by Quain (2018), which of the following should be the first step to make?

- A) Collect the information
- B) Have a good understanding of the decision you have to make
- C) Select the best alternative
- D) Evaluate the impact of your decision

✓ **Answer: B) Have a good understanding of the decision you have to make**

Which among the following is the oldest type of system used for decision support?

- A) Expert Systems
- B) Business Intelligence
- C) Enterprise Resource Planning
- D) Executive Information Systems

✓ **Answer: A) Expert Systems**

In the decision-making process, which of the following is NOT an activity of the design phase?

- A) Checking for alternatives
- B) Developing a model
- C) Setting criteria for choice
- D) Classifying the problem

✓ **Answer: D) Classifying the problem**

Which of the following is a type of control in the decision support framework?

- A) Group control
- B) Systems control
- C) Strategic control
- D) All of the above are true

✓ **Answer: C) Strategic control**

Which of the following creates a greater need for companies to use decision support systems?

- A) More companies using software as a service (SAAS)
- B) The much greater number of point-of-sale (POS) terminals
- C) Fast-changing business requirements and growing competition
- D) Greater speed and availability of the Internet

✓ **Answer: C) Fast-changing business requirements and growing competition**

What type of analytics answers questions like "what happened?"

- A) Descriptive
- B) Predictive
- C) Replication
- D) Prescriptive

✓ **Answer: A) Descriptive**

Which of the following is considered the foundation of the business intelligence architecture?

- A) ERP
- B) Data Warehousing
- C) Data Mining
- D) Big Data

✓ **Answer: B) Data Warehousing**

In the decision-making process, which of the following is NOT an activity of the design phase?

- A) Problem classification
- B) Setting criteria for choice
- C) Checking for alternatives
- D) Develop a model

✓ **Answer: A) Problem classification**

Which of the following is a type of decision in the decision support framework?

- A) Semi-structured
- B) Group decision
- C) Ad hoc
- D) All of the above are true

✓ **Answer: A) Semi-structured**

A knowledge base can be found in which of the following?

- A) Expert system
- B) Chatbot
- C) Decision support system
- D) All of the options are true

✓ **Answer: D) All of the options are true**

An integrated system where the daily functional transactions are reflected:

- A) Expert system
- B) Social network
- C) ERP
- D) None of the options are true

✓ **Answer: C) ERP**

Which of the following is NOT a main feature of a data warehouse?

- A) Reflects daily transactions
- B) Collects data from different systems
- C) Holds select historical data
- D) Enables forecasting and predictive analysis

✓ **Answer: A) Reflects daily transactions**

Which of the following can be used for forecasting future events?

- A) Data warehousing
- B) Data mining

- C) Big data
- D) All of the options are true

✓ **Answer: D) All of the options are true**

Which of the following is closely related to Business Intelligence?

- A) Data warehousing
- B) Data mining
- C) Executive information systems
- D) All of the options are true

✓ **Answer: A) Data warehousing**

What is a data mart?

- A) A subset of a data warehouse
- B) A dedicated data-cleansing system
- C) Enables loading data to the data warehouse
- D) All of the options are true

✓ **Answer: A) A subset of a data warehouse**

What makes DSS an important requirement for companies?

- A) Greater competition between companies in a fast-paced business world
- B) Greater expansion of cloud services for both storage and software as a service
- C) Greater hardware and network capabilities
- D) Greater speed and availability of the Internet

✓ **Answer: A) Greater competition between companies in a fast-paced business world**

Which is the third step in the 4-step managerial decision-making process?

- A) Create an appropriate model
- B) Determine possible solutions and evaluate them
- C) Recommend the best solution to the problem
- D) Understand the problem completely

✓ Answer: B) Determine possible solutions and evaluate them

Which of the following is a factor that may impact your company decisions?

- A) All of the options are true
- B) Change in government policies on related business regulations
- C) The great increase in overseas shipping costs
- D) The continued monthly decrease in the number of company customers

✓ Answer: A) All of the options are true

During which phase of the decision-making process is data collection conducted?

- A) Intelligence
- B) Design
- C) Choice
- D) Implementation

✓ Answer: A) Intelligence

After which phase of the decision-making process is verification and testing of the suggested solution conducted?

- A) Intelligence
- B) Choice
- C) Design
- D) Implementation

✓ Answer: B) Choice

At which level of an organization is operational control carried out?

- A) Department level
- B) Upper management level
- C) Division level
- D) All of the options are true

✓ Answer: A) Department level

Planning social responsibility and research and development are considered to be:

- A) Structured decisions
- B) Non-structured decisions
- C) Semi-structured decisions
- D) None of the options are true

✓ **Answer: B) Non-structured decisions**

A characteristic that is more associated with a DSS than typical data-based systems:

- A) Multiple users
- B) Mathematical modeling
- C) User-friendly interface
- D) Reports generation

✓ **Answer: B) Mathematical modeling**

What does ERP stand for?

- A) Enterprise Resource Planning
- B) Enterprise Records Planning
- C) Electronic Resource Planning
- D) Electronic Records Planning

✓ **Answer: A) Enterprise Resource Planning**

DSS has undergone several naming transformations. What is the latest popular name?

- A) Analytics
- B) Executive Information Systems
- C) Business Intelligence
- D) None of the above

✓ **Answer: A) Analytics**

What is a main difference between an ERP system & previous functional systems?

- A) All of the options are true
- B) It gets its data from the company's data warehouse
- C) It enables reports generation
- D) It presents a single source of truth for data items in the system

✓ **Answer: D) It presents a single source of truth for data items in the system**

Business Intelligence systems are NOT used for:

- A) Conducting analysis and producing reports based on historical data
- B) Data warehousing processes
- C) Running day-to-day business transactions
- D) Enabling business performance management

✓ **Answer: C) Running day-to-day business transactions**

Which of the following is an important source of data for a data warehouse?

- A) Co. website sales transactions
- B) ERP system
- C) Social media communications
- D) All of the options are true

✓ **Answer: D) All of the options are true**

What is the ETL process related to data warehousing used for?

- A) All of the options are true
- B) Extracting data from a data warehouse for analysis and reporting purposes
- C) Extracting data from the enterprise data warehouse and loading it into a data mart
- D) Extracting the data from ERP and other sources to be loaded into the data warehouse

✓ **Answer: D) Extracting the data from ERP and other sources to be loaded into the data warehouse**

What is the main purpose of a data mart?

- A) It is the only system where users may be able to generate analytics reports
- B) It is a staging area for data before being loaded into the data warehouse

- C) It is the only data storage where data mining models can be executed
- D) It is a subset of a data warehouse including data for a specific functional area

 **Answer: D) It is a subset of a data warehouse including data for a specific functional area**

What is the main purpose of an operational data store (ODS)?

- A) It is the only system where users may be able to generate analytics reports
- B) It is a staging area for data before being loaded into the data warehouse
- C) It is the only data storage where data mining models can be executed
- D) It is a subset of a data warehouse including data for a specific functional area

 **Answer: B) It is a staging area for data before being loaded into the data warehouse**

An enabler for descriptive analytics includes which of the following?

- A) Expert systems
- B) Forecasting
- C) Data warehousing
- D) None of the above

 **Answer: C) Data warehousing**

Predictive analytics aims to accomplish which of the following?

- A) All of the options are true
- B) Give decision makers guidance on which action to take
- C) Gives a good idea of what to expect in the future
- D) Describes clearly the current situation at hand

 **Answer: C) Gives a good idea of what to expect in the future**

In dealing with big data, the Hadoop Distributed File Systems (HDFS) enables the physical storage of data on which of the following?

- A) A single super data warehouse
- B) Many different servers
- C) All of the options are true

✓ Answer: B) Many different servers

Big data may include which of the following data?

- A) Streaming videos
- B) Web logs
- C) Image files
- D) All of the options are true

✓ Answer: D) All of the options are true

Data collection is an activity that should be completed during which stage of the decision-making process?

- A) Design
- B) Implementation
- C) Intelligence
- D) Choice

✓ Answer: C) Intelligence

Under which type of control in the decision support framework does "monitoring investment portfolio" typically fall?

- A) Strategic Planning
- B) Managerial
- C) All options are true
- D) Operational

✓ Answer: A) Strategic Planning

A main characteristic of a decision support system is:

- A) Provides models
- B) Provides data access
- C) Can be web-based
- D) All options are true

✓ Answer: D) All options are true

Making a recommendation of an action to take falls under which type of business analytics?

- A) Predictive
- B) Descriptive
- C) Prescriptive
- D) Replication

✓ **Answer: C) Prescriptive**

Which is most suitable for reflecting daily transactions and issuing daily reports?

- A) Data mart
- B) ERP
- C) Data Warehouse
- D) All options are true

✓ **Answer: B) ERP**

Which is considered the foundation of the business intelligence architecture?

- A) ERP
- B) Data Warehousing
- C) Data Mining
- D) Big Data

✓ **Answer: B) Data Warehousing**

What does OLAP stand for?

- A) Organizational Leadership Access Program
- B) Open Language Architecture Protocol
- C) On-Line Analytical Processing
- D) Object Loading App Processor

✓ **Answer: C) On-Line Analytical Processing**

Which is considered a smaller "subject-based" type of data warehouse?

- A) Enterprise data warehouse
- B) Analytics report
- C) Dashboard
- D) Data mart

✓ Answer: D) Data mart

What type of analytics answers questions like "what happened?"

- A) Descriptive
- B) Predictive
- C) Replication
- D) Prescriptive

✓ Answer: A) Descriptive

Based on the data warehouse framework as described in the course textbook, which of the following is NOT one of the known data warehousing ETL processes?

- A) Extract
- B) Execute
- C) Select
- D) Load

✓ Answer: B) Execute

Which of the following is an important tool for predictive business analytics?

- A) Data mining
- B) Dashboards
- C) Simulation
- D) All options are true

✓ Answer: A) Data mining

Which of the following is a problem with analytics?

- A) Quality of data
- B) Inaccurate data

- C) Incomplete data
- D) All options are true

✓ **Answer: D) All options are true**

Which type of data could be part of big data?

- A) Data from video recordings
- B) Data gathered from social media text messaging
- C) Data collected from sensors
- D) All options are true

✓ **Answer: D) All options are true**

What is the main objective of the MapReduce programming paradigm?

- A) Eliminate parts of big data that are irrelevant from processing
- B) Distribute big data processing over the network
- C) Distribute storage of big data over the network
- D) All options are true

✓ **Answer: B) Distribute big data processing over the network**

Which of the following is considered to be a major goal of AI?

- A) Answer questions
- B) Understand humans
- C) Do things better over time
- D) All options are true

✓ **Answer: D) All options are true**

Which of the following is a type of weak AI?

- A) Diagnosis systems
- B) Alexa
- C) Automated call center
- D) All options are true

✓ **Answer: D) All options are true**

Which of the following is among the known technologies and applications of AI?

- A) Autonomous vehicles
- B) Automatic coding
- C) Robo advisors
- D) All options are true

✓ **Answer: D) All options are true**

Chapter 2:

Which of the following is not a technology or application of AI?

- A) Speech understanding
- B) Smart homes
- C) Statistics
- D) Machine learning

✓ **Answer: C) Statistics**

Which of the following strives to have computers produce ordinary spoken language so that people can understand the computers more easily?

- A) Natural language processing
- B) Intelligent agents
- C) Natural language generation
- D) Natural language understanding

✓ **Answer: C) Natural language generation**

For many intelligent systems to work, it is necessary for them to have knowledge. The process of acquiring this knowledge is referred to as:

- A) Natural intelligence processing
- B) Knowledge acquisition
- C) Deep learning
- D) Knowledge representation

✓ **Answer: B) Knowledge acquisition**

Robots can help with online shopping by collecting shopping information, matching buyers and products, and conducting price and capability comparisons. These are known as:

- A) EC droids
- B) CRM
- C) Shopbots
- D) Turing AI toys

✓ **Answer: C) Shopbots**

Which of the following uses neural networks as part of its intelligence processing?

- A) Expert systems
- B) Knowledge systems
- C) Machine learning
- D) Deep learning

✓ **Answer: D) Deep learning**

Which of the following is NOT a biometric-related AI technology?

- A) Natural language processing
- B) Industrial robotics
- C) Image recognition
- D) Voice recognition

✓ **Answer: B) Industrial robotics**

Which of the following is NOT an AI application?

- A) Smart homes
- B) Prediction
- C) Robotics
- D) Point-of-sales transaction processing

✓ **Answer: D) Point-of-sales transaction processing**

Which of the following is an example of strong AI applications?

- A) Chess game
- B) Medical diagnosis
- C) Autonomous vehicles
- D) Automated call centers

✓ **Answer: C) Autonomous vehicles**

Monitoring systems are an example of which of the following?

- A) Assisted intelligence
- B) Autonomous AI
- C) Augmented intelligence
- D) Strong AI

✓ **Answer: A) Assisted intelligence**

Which of the following is not a main area/theory used in Analytics?

- A) Statistics
- B) Cognitive computing
- C) Operations research
- D) Management science

✓ **Answer: B) Cognitive computing**

Which of the following is one of the major AI technologies?

- A) Cloud computing
- B) Machine learning
- C) Parallel processing
- D) ERP

✓ **Answer: B) Machine learning**

Which of the following is NOT a major goal of AI?

- A) Be able to speak and respond to humans
- B) Be able to think like humans
- C) Be able to retrieve data from a database quickly
- D) Improve processing capabilities over time

✓ **Answer: C) Be able to retrieve data from a database quickly**

Which of the following is NOT one of the known AI applications?

- A) Recommender systems
- B) Sensors
- C) Language translation
- D) Autonomous vehicles

✓ **Answer: B) Sensors**

Alexa and Siri are examples of what type of AI?

- A) Knowledge-based systems
- B) Personalized services
- C) Voice recognition
- D) All options are true

✓ **Answer: D) All options are true**

Which of the following is NOT a biometric-related AI technology?

- A) Voice recognition
- B) Industrial (manufacturing) robotics
- C) Natural language processing
- D) Face recognition

✓ **Answer: B) Industrial (manufacturing) robotics**

Which of the following is NOT a major AI technology?

- A) All options given are considered major AI technologies
- B) Machine and Computer Vision
- C) Video Analytics
- D) Robotic systems

✓ Answer: A) All options given are considered major AI technologies

Which of the following AI technology is MOST likely to use a camera?

- A) Expert Systems
- B) None of the options
- C) Natural Language Processing
- D) Robotics

✓ Answer: D) Robotics

Which of the following AI technologies is likely to enable communication with people in their native language?

- A) Natural Language Processing
- B) All options are true
- C) Chatbots
- D) Robotics

✓ Answer: B) All options are true

The e-commerce "Hive" (one of the examples viewed through YouTube in the classroom) is MOST closely associated with which AI technology?

- A) Robotics
- B) Deep Learning
- C) Machine Learning
- D) Knowledge-Based Systems

✓ Answer: A) Robotics

In knowledge-based systems, the process of getting this knowledge is referred to as:

- A) Knowledge acquisition
- B) Knowledge organization and representation
- C) Knowledge refining
- D) Natural language understanding

✓ Answer: A) Knowledge acquisition

Which of the following is NOT one of the major AI technologies?

- A) Parallel processing
- B) Voice recognition
- C) Computer Vision
- D) Machine learning

✓ **Answer: A) Parallel processing**

An autonomous, small computer program that acts upon changing environments as directed by stored knowledge is called:

- A) Machine learning
- B) Intelligent agent
- C) ERP
- D) Expert system

✓ **Answer: B) Intelligent agent**

What type of system generates expert advice and gives recommendations?

- A) Machine learning
- B) ERP
- C) Expert system

✓ **Answer: C) Expert system**

Integration of digital information within the user environment in real-time, providing people real-world interactive experience with the environment is:

- A) Augmentation
- B) NLP
- C) Cognitive computing
- D) Deep Learning

✓ **Answer: A) Augmentation**

Natural Language Processing (NLP) is used for which of the following?

- A) Speech generation
- B) Voice understanding
- C) Text recognition
- D) All options are true

✓ **Answer: D) All options are true**

The ability to sit with experts and extracting (getting) their knowledge in order to develop an important DSS is called:

- A) Knowledge acquisition
- B) Knowledge collection
- C) Knowledge representation
- D) None of the options are true

✓ **Answer: A) Knowledge acquisition**

The persons who are responsible for extracting knowledge from experts are called:

- A) Knowledge developer
- B) Knowledge collector
- C) Knowledge engineer
- D) None of the options are true

✓ **Answer: C) Knowledge engineer**

Walmart has started using AI in some of its stores, which of the following is a main component in their data analysis?

- A) Big data
- B) Image recognition
- C) Video analysis
- D) All options are true

✓ **Answer: D) All options are true**

Which of the following is AI used in Human Resources management?

- A) Future sales forecasting
- B) Detecting hackers

- C) Personalized training
- D) All options are true

✓ **Answer: C) Personalized training**

Which of the following is AI used in marketing and advertising?

- A) Customer segmentation
- B) Website design
- C) All options are true
- D) Dynamic pricing

✓ **Answer: C) All options are true**

Prescriptive analytics is the only one that is considered to help decision-makers in which of the following:

- A) Determine the expected future
- B) Recommends action to take
- C) Explains the current situation
- D) All of the options are true

✓ **Answer: B) Recommends action to take**

Artificial intelligence is about which of the following:

- A) Understanding human language
- B) Reducing efforts by humans
- C) Reducing costs of performing work
- D) All of the options are true

✓ **Answer: D) All of the options are true**

Which of the following is an example of assisted intelligence:

- A) Healthcare diagnosis
- B) Self-driving cars
- C) Cybercrime fighting
- D) All of the options are true

✓ **Answer: A) Healthcare diagnosis**

What is one of the main problems of analytics:

- A) It uses mathematical modeling
- B) Used data may not be accurate
- C) It uses statistical analysis
- D) All of the options are true

✓ **Answer: B) Used data may not be accurate**

One of the main AI technologies is:

- A) Robotics
- B) Voice recognition
- C) Intelligent agents
- D) All of the options are true

✓ **Answer: D) All of the options are true**

Which of the following is best known for recommending products:

- A) Robotics
- B) Natural language processing
- C) Intelligent agents
- D) All of the options are true

✓ **Answer: C) Intelligent agents**

Which of the following may be used for retail shelf analysis:

- A) Robotics
- B) Machine and computer vision
- C) Video analytics
- D) All of the options are true

✓ **Answer: D) All of the options are true**

Natural language processing is mainly used for:

- A) Understanding written text
- B) Understanding speech
- C) Generating voice responses
- D) All of the options are true

✓ **Answer: D) All of the options are true**

A main type of AI system where extracted knowledge of people can be saved:

- A) Expert system
- B) Deep learning
- C) Intelligent agent
- D) All of the options are true

✓ **Answer: A) Expert system**

A main advantage of using AI in Human Resources:

- A) Hiring the right persons
- B) Remote access to HR data
- C) Easily applying for vacation days
- D) All of the options are true

✓ **Answer: A) Hiring the right persons**

A main advantage of using AI in marketing:

- A) Predicting pricing of products
- B) Improved search engine results
- C) Targeting ads to the right people
- D) All of the options are true

✓ **Answer: D) All of the options are true**

Natural language processing (NLP) involves which of the following technologies:

- A) Speech recognition
- B) Image recognition

- C) Machine and computer vision
- D) All options are true

✓ **Answer: A) Speech recognition**

Computers that learn from examples, big data, and new situations while making use of data collected from sensors and other resources is what type of technology?

- A) Machine learning
- B) Expert systems
- C) Forecasting
- D) Data Warehousing

✓ **Answer: A) Machine learning**

The process of observing, interviewing, and discussing falls under which step of knowledge-based systems creation?

- A) Reasoning from knowledge
- B) Knowledge acquisition
- C) Knowledge representation
- D) All options are true

✓ **Answer: B) Knowledge acquisition**

Which is considered the most difficult task of building a knowledge-based system?

- A) Explanation justification
- B) Knowledge acquisition
- C) Natural language generation
- D) Response generation

✓ **Answer: B) Knowledge acquisition**

The integration of digital information within the user environment in real-time, providing people real-world interactive experience with the environment is called:

- A) Image recognition
- B) Augmented reality

- C) Knowledge acquisition
- D) Integration agents

✓ **Answer: B) Augmented reality**

Analytics is most known for using the following, EXCEPT:

- A) Statistics
- B) Mathematical modeling
- C) Knowledge
- D) Big data

✓ **Answer: C) Knowledge**

All of the following are types of AI applications, EXCEPT:

- A) Autonomous cars
- B) Smart homes
- C) Robotics
- D) Point-of-sale systems

✓ **Answer: D) Point-of-sale systems**

Predictive maintenance is possible through which of the following AI technologies:

- A) NLP
- B) Video analytics
- C) Knowledge-based systems
- D) Machine learning

✓ **Answer: D) Machine learning**

Neural networks are mainly associated with which of the following AI technologies:

- A) Machine and computer vision
- B) Video analytics
- C) Expert systems
- D) Deep learning

✓ **Answer: D) Deep learning**

Complex problems that cannot be dealt with using machine learning are solved by:

- A) Deep learning
- B) Video analytics
- C) Expert systems
- D) Machine and computer vision

✓ **Answer: A) Deep learning**

The "object counter" application is an example of which AI technology:

- A) Chatbots
- B) NLP
- C) Expert systems
- D) Machine and computer vision

✓ **Answer: D) Machine and computer vision**

One of the capabilities of a chatbot is:

- A) Understanding written text
- B) Translating languages
- C) Understanding humans
- D) All options are true

✓ **Answer: D) All options are true**

Chapter 4:

How can data mining help in reducing death related to cancer?

- A) By allowing doctors to ask questions about the patients.
- B) By learning about previous effective treatments under different circumstances.
- C) By allowing easy tracking of the patient history that is being treated.
- D) All options are true.

✓ **Answer: B) By learning about previous effective treatments under different circumstances.**

Which of the following algorithms is the most used for segmentation?

- A) Decision trees.
- B) Apriori.
- C) K-means.
- D) K-fold cross-validation.

✓ **Answer: C) K-means.**

Which of the following is an approach to determine which data mining estimation methodology produced more accurate prediction?

- A) Single split.
- B) K-fold cross-validation.
- C) Confusion matrix.
- D) All options are true.

✓ **Answer: C) Confusion matrix.**

If k=5 in a K-fold cross-validation methodology, what percent of the data is used as training data?

- A) 10%.
- B) 80%.
- C) 20%.
- D) 5%.

✓ **Answer: B) 80%.**

Which of the following is an example of nominal data?

- A) White, green, blue.
- B) High, middle, low.
- C) First, second, third.
- D) All options are true.

✓ **Answer: A) White, green, blue.**

How did the main champion of the analytics system in the Miami-Dade County Police Department convince detectives to accept the idea of using an analytics system for fighting crime?

- A) Showed everyone a YouTube video on the importance of Analytics.
- B) Explain to them how the system works.
- C) Use the system to solve an open unsolved case.
- D) Had the young new more IT savvy detectives convince the older detectives.

✓ **Answer: C) Use the system to solve an open unsolved case.**

Which of the following can data mining be used with:

- A) Unstructured data.
- B) Data warehouses.
- C) Databases.
- D) All options are true.

✓ **Answer: D) All options are true.**

Which broad area of data mining applications divides a collection of objects into natural groupings with similar features?

- A) Clustering.
- B) Associations.
- C) Visualization.
- D) Classification.

✓ **Answer: A) Clustering.**

Supervised learning is used mainly with which data mining tasks:

- A) Segmentation.
- B) Association.
- C) Prediction.
- D) All options are true.

✓ **Answer: C) Prediction.**

In data mining, finding a relationship between two products to be commonly together in a shopping cart is known as:

- A) Artificial neural networks.
- B) Association rule mining.
- C) Cluster analysis.
- D) Decision trees.

✓ **Answer: B) Association rule mining.**

Which of the following areas can be considered part of data mining?

- A) Management Science.
- B) Artificial Intelligence.
- C) Information visualization.
- D) All options are true.

✓ **Answer: D) All options are true.**

What was/were the event(s) that helped the Miami-Dade police department accept data mining in their crime detection operations?

- A) A much higher crime rate than the last year.
- B) The recommendation by the state government to start using technology in crime fighting.
- C) The solving of some previously unsolved crimes.
- D) All options are true.

✓ **Answer: C) The solving of some previously unsolved crimes.**

Which of the following is/are important source(s) of data for data mining?

- A) A data warehouse.
- B) Web log files.
- C) System databases.
- D) All options are true.

✓ **Answer: D) All options are true.**

In association types of patterns:

- A) Future expected events or values are determined.
- B) The natural grouping of things based on their characteristics.
- C) The grouping of related things.
- D) The determination of things to come in order.

✓ **Answer: C) The grouping of related things.**

In segmentation types of patterns:

- A) Future expected events or values are determined.
- B) The natural grouping of things based on their characteristics.
- C) The grouping of related things.
- D) The determination of things to come in order.

✓ **Answer: B) The natural grouping of things based on their characteristics.**

Which of the following decrease the size and complexity of problems for other data mining methods?

- A) Association mining.
- B) Apriori algorithm.
- C) Classification.
- D) Cluster analysis.

✓ **Answer: D) Cluster analysis.**

When using the K-means clustering algorithm, what is the expected value of K?

- A) Always 4.
- B) Always 5.
- C) Depends on the number of clusters we would like to create.
- D) Depends on the number of attributes that are used as inputs.

✓ **Answer: C) Depends on the number of clusters we would like to create.**

What are the typical outputs of running an Apriori algorithm?

- A) Number of repetition of two or more items in transaction records.
- B) Number of clusters in the data.

- C) Raw transaction data.
- D) All options are true.

✓ **Answer: A) Number of repetition of two or more items in transaction records.**

How is text mining different from data mining?

- A) Text mining is a semi-automated process.
- B) Text mining works with unstructured data.
- C) Text mining searches for patterns.
- D) All options are true.

✓ **Answer: B) Text mining works with unstructured data.**

What is the purpose of stemming?

- A) Determining the number of repeated words.
- B) Determining the root of words.
- C) Creating a decision tree structure.
- D) Removing words like "a", "an", "to", etc.

✓ **Answer: B) Determining the root of words.**

Which of the data mining processes involves a step that tries to get a good idea about the objective of the desired data mining model to be developed?

- A) Classification.
- B) CRISP-DM.
- C) SEMMA.
- D) KDD.

✓ **Answer: B) CRISP-DM.**

Which of the data mining processes is considered to be a more comprehensive process which actually involves a data-mining step:

- A) Classification.
- B) CRISP-DM.
- C) SEMMA.
- D) KDD.

✓ Answer: D) KDD.

In the CRISP-DM data mining process, the phase that takes the most of the project time is:

- A) Data Understanding.
- B) Data Preparation.
- C) Model Building.
- D) Testing and Evaluation.

✓ Answer: B) Data Preparation.

The most used data mining method is:

- A) Clustering.
- B) Classification.
- C) Time-series.
- D) Regression.

✓ Answer: B) Classification.

Which of the following is NOT an example of nominal data:

- A) High, middle, low.
- B) Open, closed.
- C) Blue, green, yellow.
- D) Pen, pencil, marker.

✓ Answer: A) High, middle, low.

According to the taxonomy of data mining as defined in the textbook, which of the following is NOT a data mining task:

- A) Prediction.
- B) Segmentation.
- C) Association.
- D) Link analysis.

✓ Answer: D) Link analysis.

According to the taxonomy of data mining as defined in the textbook, which of the following is NOT a data mining method:

- A) Regression.
- B) Knowledge discovery.
- C) Clustering.
- D) Market-basket.

✓ **Answer: B) Knowledge discovery.**

When it comes to types of learning in a data mining model, unsupervised learning involves which of the following:

- A) Providing only output values to the model.
- B) Providing both input and output values to the model.
- C) Providing only input values to the model.
- D) None of the options is true.

✓ **Answer: C) Providing only input values to the model.**

Which data mining method is used to determine the future expected value of the same variable:

- A) Link analysis.
- B) Market-basket.
- C) Sequence analysis.
- D) Time series.

✓ **Answer: D) Time series.**

Which data mining method is used to discover relationships between groups of people or things:

- A) Link analysis.
- B) Market-basket.
- C) Sequence analysis.
- D) Time series.

✓ **Answer: A) Link analysis.**

Supervised learning is used mainly with which data mining tasks?

- A) Association.
- B) Prediction.
- C) Clustering.
- D) Segmentation.

✓ **Answer: B) Prediction.**

In data mining, finding a relationship between two or more persons on a social media platform, or two or more websites would be done through:

- A) Market-basket analysis.
- B) Link analysis.
- C) Sequence analysis.
- D) Segmentation.

✓ **Answer: B) Link analysis.**

Which of the following is the first step of the CRISP-DM process?

- A) Data understanding.
- B) Data preparation.
- C) Business understanding.
- D) Model building.

✓ **Answer: C) Business understanding.**

In the simple split test estimation methodology, what percent of the data should be used for training the model?

- A) 1/2.
- B) 1/3.
- C) 1/4.
- D) 2/3.

✓ **Answer: D) 2/3.**

Which of the following estimation methodology for classification produces more accurate prediction?

- A) Single split.
- B) K-fold cross-validation.
- C) Confusion matrix.
- D) All options are true.

✓ **Answer: B) K-fold cross-validation.**

What is a different name for single split estimation methodology?

- A) Simple split.
- B) Holdout.
- C) Sample test estimation.
- D) All options are true.

✓ **Answer: D) All options are true.**

In single split estimation methodology, what percent of the data is used for training the model?

- A) 1/2.
- B) 1/3.
- C) 1/4.
- D) 2/3.

✓ **Answer: D) 2/3.**

If k=5 in K-fold cross-validation methodology, what percent of the data is used as testing data in each run?

- A) 10%.
- B) 80%.
- C) 20%.
- D) 5%.

✓ **Answer: C) 20%.**

What is the main purpose of using a confusion matrix?

- A) Determine which data to use for training and testing the model.
- B) Determine the accuracy of the developed classification model.
- C) Determine if there is error or missing data.
- D) All options are true.

✓ **Answer: B) Determine the accuracy of the developed classification model.**

One of the main differences between data mining and statistics?

- A) Data mining is used to help in decision-making while statistics is for knowing some information about populations.
- B) Data mining works better with big data while statistics only requires some sample of the data from the general population.
- C) Data mining can use computerized systems while statistics are always solved manually.
- D) All options are true.

✓ **Answer: B) Data mining works better with big data while statistics only requires some sample of the data from the general population.**

In the area of customer relationship management, data mining can be valuable for which of the following?

- A) Having an improved, more efficient email server.
- B) Tracking the most sold and most popular products.
- C) Targeting email messages at different customers based on their level of satisfaction with company services.
- D) All options are true.

✓ **Answer: C) Targeting email messages at different customers based on their level of satisfaction with company services.**

In the area of retailing and logistics, data mining can be used for which of the following?

- A) Make sure stores are stocked with an appropriate number of specific items.
- B) Make sure the company is ready for customer delivery demands based on the time of the year.
- C) Arrange products within the store in a way to help better selling of these products.
- D) All options are true.

✓ Answer: D) All options are true.

One of the most popular data mining processes:

- A) SEMMA.
- B) KDD.
- C) CRISP-DM.
- D) All options are true.

✓ Answer: C) CRISP-DM.

Which of the following data mining methods is more useful for finding relationships between people?

- A) Regression.
- B) Classification.
- C) Link analysis.
- D) Sequence analysis.

✓ Answer: C) Link analysis.

How did the main champion of the analytics system in the Miami-Dade County Police Department convince detectives to accept the idea of using an analytics system for fighting crime?

- A) Showed everyone a YouTube video on the importance of analytics.
- B) Use the system to solve an open unsolved case.
- C) Had the young new more IT-savvy detectives convince the older detectives.
- D) All options are true.

✓ Answer: B) Use the system to solve an open unsolved case.

What of the following factors makes using data mining important for businesses?

- A) The ability to consolidate historical and new data into one data source.
- B) Data becoming one of the most important assets for an organization.
- C) Increased data processing and storage capabilities and decreased costs.
- D) All options are true.

✓ Answer: D) All options are true.

Which broad area of data mining applications divides a collection of objects into natural groupings with similar features?

- A) Associations.
- B) Visualization.
- C) Classification.
- D) Clustering.

✓ Answer: D) Clustering.

Supervised learning is used mainly with which data mining tasks?

- A) Association.
- B) Prediction.
- C) Segmentation.
- D) Visualization.

✓ Answer: B) Prediction.

In data mining, finding a relationship between two products to be commonly found together, for example, as in a shopping cart, is known as:

- A) Association rule mining.
- B) Classification.
- C) Decision trees.
- D) Time-series analysis.

✓ Answer: A) Association rule mining.

In decision trees, "pruning" is concerned with which of the following:

- A) Determining when to stop splitting the decision tree.
- B) The variable and value that is used for determining how to split the decision tree.
- C) Determining which part of the decision tree leaf nodes to remove.
- D) All options are true.

✓ Answer: C) Determining which part of the decision tree leaf nodes to remove.

After which of the following stages will a decision tree be able to give output results?

- A) The created model has been trained.
- B) The classification algorithm has been selected.
- C) The classification rules have been generated.
- D) All options are true.

✓ **Answer: D) All options are true.**

The K-means algorithm used for clustering will reach its conclusion when:

- A) The center cluster points values no longer change.
- B) It has moved the cluster center point K times.
- C) K number of iterations of the algorithm have been executed.
- D) All options are true.

✓ **Answer: A) The center cluster points values no longer change.**

What does the $\text{Supp}(X \rightarrow Y)$ function in association rule mining give?

- A) The number of times both Z and Y appear together.
- B) The percentage of X + Y from the total number of transactions available.
- C) The percentage of X from the total number of X=Y.
- D) The percentage of Y from the total number of X=Y.

✓ **Answer: B) The percentage of X + Y from the total number of transactions available.**

The minimum support value in association rule mining is used for:

- A) The minimum number of iterations the algorithm must run.
- B) The minimum number of associations that the algorithm must find.
- C) The minimum number of products in a basket for it to be considered for association.
- D) The minimum acceptable number for a confidence value.

✓ **Answer: C) The minimum number of products in a basket for it to be considered for association.**

The stage in the CRISP-DM data mining process where the developer checks to see if data necessary for developing the data mining model exists is:

- A) Business Understanding.
- B) Data Understanding.
- C) Data Preparation.
- D) Testing & Evaluation.

✓ **Answer: B) Data Understanding.**

What is the purpose of data transformation in the KDD data mining process?

- A) Make sure the data is cleansed.
- B) Select the most appropriate data for the model.
- C) Make changes where needed to the data so that it fits better for the model.
- D) All options are true.

✓ **Answer: C) Make changes where needed to the data so that it fits better for the model.**

The most appropriate algorithm used with market basket analysis is:

- A) Apriori algorithm.
- B) Decision trees.
- C) K-means algorithm.
- D) Support Vector Machines algorithm.

✓ **Answer: A) Apriori algorithm.**

In the simple split estimation method for classification, what is the typical percentage of data used for training the model:

- A) 1/4.
- B) 1/3.
- C) 1/2.
- D) 2/3.

✓ **Answer: D) 2/3.**

In an 8-fold cross-validation estimation method for classification, what is the typical percentage of data used for testing the model:

- A) 1/8.
- B) 1/4.
- C) 1/2.
- D) 7/8.

✓ **Answer: A) 1/8.**

Data mining is associated with which of the following?

- A) Statistics
- B) Management Science
- C) Artificial intelligence
- D) All of the options are true

✓ **Answer: D) All of the options are true**

Which of the following is NOT one of the other known names for data mining?

- A) Knowledge discovery
- B) Data dredging
- C) Information harvesting
- D) Knowledge understanding

✓ **Answer: D) Knowledge understanding**

How did data mining help the Miami-Dade police department?

- A) Predict areas with potential increase in crime
- B) Deal with shortage in funding
- C) Solve previously unsolved crimes
- D) All options are true

✓ **Answer: D) All options are true**

Clustering is a data mining method which falls under which main data mining task?

- A) Prediction
- B) Segmentation
- C) Association
- D) None of the above is true

✓ **Answer: B) Segmentation**

Sequence analysis is a data mining method which falls under which main data mining task?

- A) Prediction
- B) Segmentation
- C) Association
- D) None of the above is true

✓ **Answer: C) Association**

Supervised learning is used with which data mining methods?

- A) Regression
- B) Clustering
- C) Link analysis
- D) All options are true

✓ **Answer: A) Regression**

Which of the following is a possible application of data mining?

- A) Predict machine failure
- B) Maximize customer value
- C) Classes
- D) All options are true

✓ **Answer: D) All options are true**

Deciding on what questions we want to answer when planning to use data mining is performed during which step of the data mining process?

- A) Data understanding
- B) Business understanding

- C) Model building
- D) All options are true

✓ **Answer: B) Business understanding**

In the SEMMA data mining process, which is the last step in the process?

- A) Explore
- B) Assess
- C) Deployment
- D) Modify

✓ **Answer: B) Assess**

Recursively dividing a training set until each division consists entirely or primarily of examples from one class is conducted by which type of popular algorithm?

- A) Decision tree analysis
- B) Logistic regression
- C) K-means clustering
- D) Apriori algorithm

✓ **Answer: A) Decision tree analysis**

The main purpose of data mining is to use large sets of data to recognize:

- A) Experts in a field
- B) Patterns in the data
- C) Statistical models
- D) None of the above

✓ **Answer: B) Patterns in the data**

How did data mining help the Miami Dade police department with unsolved cases?

- A) Show videos of criminals as they conduct the crime
- B) Encourage people to give information to the police
- C) Match collected clues with previous crimes
- D) All options are true

✓ **Answer: C) Match collected clues with previous crimes**

What made data mining a much more important option these days for businesses?

- A) Greater use of data warehouses
- B) More data available
- C) Lower data processing costs
- D) All options are true

✓ **Answer: D) All options are true**

The ability to find some relationship between different entities in DM is called?

- A) Clustering
- B) Prediction
- C) Association
- D) Segmentation

✓ **Answer: C) Association**

The ability to predict the expected value of a specific entity after some time is:

- A) Classification
- B) Market-basket analysis
- C) Sequence analysis
- D) Time-series

✓ **Answer: D) Time-series**

Which is the most commonly used data-mining process by companies?

- A) CRISP-DM
- B) SEMMA
- C) KDD
- D) All are equal in popularity

✓ **Answer: A) CRISP-DM**

The step where a DM model developer needs to check if the existing data can help in creating the desired model:

- A) Data preparation
- B) Testing and verification
- C) Data understanding
- D) Business understanding

✓ **Answer: C) Data understanding**

When using the K-means clustering algorithm, where K = 4, what is the percent of training data for each iteration of the run?

- A) 4%
- B) 40%
- C) 25%
- D) 75%

✓ **Answer: D) 75%**

Which formula is used to tell us what is the percent of truly predicted classes from the overall true predictions?

- A) Precision
- B) Accuracy
- C) Recall
- D) True positive rate

✓ **Answer: A) Precision**

In decision tree analysis, how is the best splitting attribute selected to start at the root node?

- A) The first attribute in the table of selected data attributes
- B) The first nominal data attribute in the list of attributes
- C) It depends on the actual decision tree algorithm being used
- D) The first ordinal data attribute in the list of attributes

✓ **Answer: C) It depends on the actual decision tree algorithm being used**

An algorithm which is used for market basket analysis is called:

- A) Decision tree analysis
- B) Logistic regression
- C) K-means clustering
- D) Apriori algorithm

✓ **Answer: D) Apriori algorithm**

Which of the following is NOT a source(s) of data for data mining:

- A) Social networks data
- B) Company databases
- C) Company website log files
- D) All options are important sources

✓ **Answer: D) All options are important sources**

When using the K-means clustering algorithm, what is the value of K?

- A) Depends on the number of clusters we would like to create.
- B) Depends on the number of attributes that we use as inputs.
- C) Depends on the graph we use: 2 or 3 dimensional.
- D) Always 5.

✓ **Answer: A) Depends on the number of clusters we would like to create.**

When it comes to types of learning in a data mining model, supervised learning involves which of the following:

- A) Providing only output values to the model.
- B) Providing both input and output values to the model.
- C) Providing only input values to the model.
- D) None of the options is true.

✓ **Answer: B) Providing both input and output values to the model.**

Which data mining method is used to discover relationships between groups of websites:

- A) Link analysis.
- B) Market-basket.

- C) Sequence analysis.
- D) Time series.

✓ **Answer: A) Link analysis.**

In the KDD data mining process, the first phase is:

- A) Data Understanding
- B) Data Selection
- C) Data Cleaning
- D) Data Transformation

✓ **Answer: B) Data Selection**

Which of the following is an example of nominal data:

- A) High, middle, low.
- B) A grade, B grade, C grade
- C) Pen, pencil, marker.
- D) 50, 100.

✓ **Answer: C) Pen, pencil, marker.**

Which of the following is used to measure the accuracy level of prediction method?

- A) Single split.
- B) K-fold cross-validation.
- C) Confusion matrix.
- D) None of the options is true.

✓ **Answer: C) Confusion matrix.**

In single test estimation method, what percent of the data is used for testing the model:

- A) About 50%.
- B) About 30%.
- C) About 25%.
- D) About 70%.

✓ Answer: B) About 30%.

If k=4 in a K-fold cross-validation methodology, what percent of the data is used as training data in each run:

- A) 40%.
- B) 75%.
- C) 25%.
- D) 4%.

✓ Answer: B) 75%.

Chapter 7:

Sentiment analysis is most useful for which of the following types of applications:

- A) ERP.
- B) Shopping carts.
- C) CRM.
- D) Expert systems.

✓ Answer: C) CRM.

Which of the following is a step in the sentiment analysis process:

- A) Determining who or what is the target of the sentiment.
- B) Determine the objectivity of the word or statement being analyzed.
- C) Determine the sentiment of the entire analyzed document.
- D) All options are true.

✓ Answer: D) All options are true.

What does Web content mining involve:

- A) Analyzing the universal resource locators (URLs) in Web pages.
- B) Analyzing the unstructured content of Web pages.
- C) Analyzing the pattern of visits to a Web site.
- D) Analyzing the metadata of a Web page.

✓ Answer: B) Analyzing the unstructured content of Web pages.

What does Web structure mining involve:

- A) Analyzing the universal resource locators (URLs) in Web pages.
- B) Analyzing the unstructured content of Web pages.
- C) Analyzing the pattern of visits to a Web site.
- D) Analyzing the metadata of a Web page.

✓ Answer: A) Analyzing the universal resource locators (URLs) in Web pages.

After knowledge is extracted using the Web usage mining clickstream analysis, what is the potential added value:

- A) Improving the data.
- B) Improving the website design.
- C) Increasing customer value.
- D) All options are true.

✓ Answer: D) All options are true.

End of collection

Due to the lack of collections of the seventh chapter, what is below are additional questions generated using AI for the purpose of practice.

Text mining is best described as:

- A) Extracting structured knowledge from unstructured data
- B) A semi-automated process of extracting knowledge from unstructured data
- C) Mining structured databases
- D) Automating knowledge discovery in structured databases

✓ Answer: B) A semi-automated process of extracting knowledge from unstructured data

Which of the following distinguishes text mining from data mining?

- A) Text mining deals with structured data only
- B) Text mining deals with data from relational databases
- C) Text mining handles unstructured or semi-structured data
- D) Data mining focuses on unstructured data

✓ **Answer: C) Text mining handles unstructured or semi-structured data**

Which of these is NOT an application area for text mining?

- A) Summarization
- B) Categorization
- C) Webpage crawling
- D) Topic tracking

✓ **Answer: C) Webpage crawling**

What is the process of reducing inflected words to their base form?

- A) Tokenization
- B) Noise filtering
- C) Stemming
- D) Clustering

✓ **Answer: C) Stemming**

Which term refers to words like “a,” “an,” and “the” in text mining?

- A) Synonyms
- B) Stop words
- C) Polysemes
- D) Key terms

✓ **Answer: B) Stop words**

Sentiment analysis primarily aims to:

- A) Determine customer spending
- B) Detect favorable and unfavorable opinions

- C) Analyze text for grammar correctness
- D) Filter out unnecessary text

✓ Answer: B) Detect favorable and unfavorable opinions

What is the first step in the text mining process?

- A) Establish the corpus
- B) Create a term-document matrix
- C) Generate predictive patterns
- D) Apply clustering algorithms

✓ Answer: A) Establish the corpus

The term ‘corpus’ in text mining refers to:

- A) A large collection of structured or unstructured texts
- B) A synonym database
- C) A key term extraction algorithm
- D) The output of NLP methods

✓ Answer: A) A large collection of structured or unstructured texts

Which of the following involves grouping similar documents without predefined categories?

- A) Categorization
- B) Clustering
- C) Summarization
- D) Concept linking

✓ Answer: B) Clustering

Which task in sentiment analysis identifies whether text is objective or subjective?

- A) Target identification
- B) Sentiment detection
- C) Polarity classification
- D) Aggregation

✓ Answer: B) Sentiment detection

Which component of a search engine systematically browses the web for new pages?

- A) Query analyzer
- B) Web crawler
- C) Document ranker
- D) Metadata tagger

✓ Answer: B) Web crawler

What does Web content mining focus on?

- A) Extracting information from unstructured web page content
- B) Analyzing hyperlink patterns
- C) Analyzing clickstream data
- D) Ranking web pages

✓ Answer: A) Extracting information from unstructured web page content

In sentiment analysis, what does “P-N polarity” refer to?

- A) Positive versus Negative sentiment
- B) Neutral versus Passive sentiment
- C) Pattern versus Noise classification
- D) Primary versus Non-primary sentiment

✓ Answer: A) Positive versus Negative sentiment

What does WordNet NOT provide?

- A) Synonyms and antonyms
- B) Semantic relations between words
- C) Automated clustering algorithms
- D) Hand-coded word definitions

✓ Answer: C) Automated clustering algorithms

Which of the following is NOT part of Natural Language Processing (NLP)?

- A) Machine translation
- B) Speech recognition
- C) Optical character recognition
- D) Video analysis

✓ **Answer: D) Video analysis**

What are the three main categories of Web mining?

- A) Content mining, Structure mining, Usage mining
- B) Data mining, Text mining, Usage mining
- C) Semantic mining, Structure mining, Content mining
- D) Text mining, Clickstream analysis, Content mining

✓ **Answer: A) Content mining, Structure mining, Usage mining**

Which method is used to create summaries of text documents in text mining?

- A) Categorization
- B) Summarization
- C) Concept linking
- D) Clustering

✓ **Answer: B) Summarization**

What challenge does web usage mining face due to web size and complexity?

- A) The web is too big and dynamic
- B) Insufficient data available
- C) Lack of proper tools
- D) Limited user access

✓ **Answer: A) The web is too big and dynamic**

Which step of text mining involves creating a term-document matrix?

- A) Establishing the corpus
- B) Structuring the data

- C) Extracting patterns
- D) Applying clustering

✓ Answer: B) Structuring the data

What is a major challenge of Natural Language Processing?

- A) Finding synonyms
- B) Handling the ambiguity of spoken language
- C) Speed of processing text
- D) Creating term-document matrices

✓ Answer: B) Handling the ambiguity of spoken language

Web structure mining focuses on:

- A) Analyzing hyperlinks between web pages
- B) Extracting text content
- C) Categorizing documents
- D) Understanding clickstream patterns

✓ Answer: A) Analyzing hyperlinks between web pages

Which algorithm rates web pages using hyperlink information?

- A) HITS algorithm
- B) CRISP-DM
- C) Apriori algorithm
- D) TF-IDF

✓ Answer: A) HITS algorithm

What is sentiment aggregation in sentiment analysis?

- A) Breaking down text into individual words
- B) Summarizing sentiment from word to document level
- C) Classifying text as subjective or objective
- D) Detecting web usage patterns

✓ Answer: B) Summarizing sentiment from word to document level

What is tokenization in text mining?

- A) Categorizing blocks of text into meaningful pieces
- B) Summarizing text documents
- C) Analyzing the frequency of terms
- D) Matching patterns in text

✓ **Answer: A) Categorizing blocks of text into meaningful pieces**

Which method reduces the dimensionality of term-document matrices?

- A) Term clustering
- B) Singular Value Decomposition (SVD)
- C) Keyword indexing
- D) Synonym matching

✓ **Answer: B) Singular Value Decomposition (SVD)**

What is the primary goal of text mining?

- A) Improve data quality
- B) Extract useful patterns from unstructured text
- C) Analyze structured data
- D) Summarize numeric data

✓ **Answer: B) Extract useful patterns from unstructured text**

Which of the following is NOT an example of text mining applications?

- A) Document summarization
- B) Spam email detection
- C) Sentiment analysis
- D) Web server monitoring

✓ **Answer: D) Web server monitoring**

What is Term Frequency-Inverse Document Frequency (TF-IDF) used for in text mining?

- A) To weigh the importance of terms in a document
- B) To cluster documents
- C) To tokenize text
- D) To visualize word relationships

✓ Answer: A) To weigh the importance of terms in a document

Which of the following is a critical task of natural language processing (NLP)?

- A) Network mining
- B) Data cleaning
- C) Part-of-speech tagging
- D) Creating web graphs

✓ Answer: C) Part-of-speech tagging

Which component of text mining is responsible for converting text into numerical data?

- A) Corpus construction
- B) Text vectorization
- C) Sentiment analysis
- D) Summarization

✓ Answer: B) Text vectorization

Which task in text mining involves identifying relationships between terms in a text?

- A) Tokenization
- B) Summarization
- C) Concept linking
- D) Categorization

✓ Answer: C) Concept linking

In text mining, what is a “bag-of-words” model?

- A) A representation of text that ignores grammar and word order
- B) A method for clustering documents

- C) A type of semantic analysis
- D) A graph-based web analysis technique

✓ Answer: A) A representation of text that ignores grammar and word order

Which of the following is a challenge in text mining?

- A) Extracting structured data
- B) Handling language ambiguity
- C) Cleaning relational databases
- D) Reducing computational costs

✓ Answer: B) Handling language ambiguity

What does Latent Semantic Analysis (LSA) aim to achieve in text mining?

- A) Sentence-level translation
- B) Tokenization of words
- C) Discovering hidden relationships between terms
- D) Clustering of similar documents

✓ Answer: C) Discovering hidden relationships between terms

What is the primary use of word embedding techniques in text mining?

- A) Clustering text data
- B) Summarizing documents
- C) Representing words in vector space
- D) Identifying synonyms

✓ Answer: C) Representing words in vector space

Which web mining category involves analyzing user behavior on websites?

- A) Web content mining
- B) Web structure mining
- C) Web usage mining
- D) Sentiment mining

✓ Answer: C) Web usage mining

Which of the following methods is most used in document clustering?

- A) Sentiment analysis
- B) K-means algorithm
- C) Apriori algorithm
- D) TF-IDF

✓ **Answer: B) K-means algorithm**

Which task in text mining involves determining if a document belongs to a specific category?

- A) Categorization
- B) Tokenization
- C) Clustering
- D) Summarization

✓ **Answer: A) Categorization**

What is the main challenge of using stop words in text analysis?

- A) They may remove important terms in specific contexts
- B) They increase term frequency
- C) They add noise to the analysis
- D) They make text too complex

✓ **Answer: A) They may remove important terms in specific contexts**

What is Named Entity Recognition (NER) in NLP?

- A) Tokenizing text into sentences
- B) Identifying proper nouns and phrases
- C) Categorizing documents
- D) Summarizing web content

✓ **Answer: B) Identifying proper nouns and phrases**

Which of the following is NOT an application of web usage mining?

- A) Clickstream analysis
- B) Recommendation systems
- C) Data encryption
- D) Customer segmentation

✓ Answer: C) Data encryption

In sentiment analysis, what does polarity detection determine?

- A) The length of the text
- B) The positivity or negativity of the text
- C) The grammatical accuracy of the text
- D) The source of the text

✓ Answer: B) The positivity or negativity of the text

Which of these techniques is used for document summarization?

- A) Apriori algorithm
- B) CRISP-DM
- C) TextRank
- D) TF-IDF

✓ Answer: C) TextRank

What is the purpose of the lemmatization process in text mining?

- A) Remove punctuation
- B) Break sentences into tokens
- C) Convert words to their base forms
- D) Rank keywords by importance

✓ Answer: C) Convert words to their base forms

What does the term "document frequency" refer to in text mining?

- A) Number of tokens in a document
- B) The average word length in a text

- C) Number of documents in which a term appears
- D) Sentiment score of a document

✓ **Answer: C) Number of documents in which a term appears**