INDIAN SCHOOL AL WADI AL KABIR

Woksheet No:1	ARTIFICIAL INTELLIGENCE(417)- CHAPTER 4 :DATA SCIENCE

1 Mark questions

1. **What was one of the earliest applications of data science?**
   A. Genetics & Genomics
   B. Internet Search
   C. Fraud and Risk Detection
   D. Targeted Advertising

2. **Which industry uses data science to predict flight delays and drive customer loyalty programs?**
   A. Genetics & Genomics
   B. Internet Search
   C. Fraud and Risk Detection
   D. Airline Route Planning

3. **Which search engine(s) make use of data science algorithms to deliver search results?**
   A. Google
   B. Yahoo
   C. Bing
   D. All of the above

4. **What is the purpose of website recommendations?**
   A. To improve user experience
   B. To promote products
   C. To suggest similar products
   D. All of the above

5. **Which field does data science enable an advanced level of treatment personalization through research in?**
   A. Genetics & Genomics
   B. Internet Search
   C. Fraud and Risk Detection
   D. Airline Route Planning

6. **Which industry uses data science algorithms to decide on digital advertising placements?**
   A. Genetics & Genomics
   B. Internet Search
   C. Targeted Advertising
   D. Airline Route Planning

7. **What is the main goal of data science applications in genetics and genomics?**
   A. To understand the impact of DNA on health
   B. To find individual biological connections
   C. To analyze reactions to drugs and diseases
   D. All of the above

8.**Which industry uses data science to analyze customer profiling and past expenditures?**
   A. Genetics & Genomics
   B. Fraud and Risk Detection
   C. Internet Search
   D. Airline Route Planning

9. **Which companies use data science to improve user experience through product recommendations?**
   A. **Amazon, Twitter, Google Play, Netflix, LinkedIn, IMDB**
   B. Yahoo, Bing, Ask, AOL
   C. All of the above
   D. None of the above

10. **What is the main purpose of data science?**
    A. **To analyze data and make machines intelligent**
    B. To collect and store data for future use
    C. To create new fields of study in science
    D. To predict future trends based on historical data

Answer: A) To analyze data and make machines intelligent

Data science involves analyzing data to extract knowledge and insights. These insights can then be used in various applications, including making machines intelligent through machine learning and artificial intelligence. While the other options involve aspects of data science, option A most accurately captures its main purpose.

11. **In which industry were the earliest applications of data science seen?**
    A. Technology
    B. Healthcare
    C. Finance
    D. Agriculture

Answer C) Finance industry: The finance industry had a lot of data and faced challenges such as losses and complications. They brought in data science practices to help solve their problems.

12. **How did data science help banking companies in managing risk?**
    A. By collecting customer profiles and past expenditures
    B. By creating new banking products
    C. By providing loans to customers with high purchasing power
    D. **By reducing bad debts and losses**

Answer D) Reducing bad debts and losses:

Banks use data science to manage various types of risks, such as credit risk, market risk, and operational risk

13. **What does genetics and genomics research aim to understand?**
    A. The impact of DNA on our health
    B. The impact of environment on our health
    C. The relationship between genes and diseases only
    D. The relationship between genes, diseases, and drug response

Answer: Genetics and genomics research aims to understand:
   1) The impact of DNA on our health
   2) The impact of environment on our health
   3) The relationship between genes and diseases only
   4) The relationship between genes, diseases, and drug response

Genetics and genomics research is a vast field that aims to understand the impact of our DNA (genes) on our health, the relationship between genes and diseases, and how these relationships influence our response to drugs. While it does consider the impact of the environment on our health, its primary focus is on our genetic makeup. So, while all the options are related to genetics and genomics research, option D most accurately captures its main aim.

14. **How does data science contribute to disease research?**
    A. It provides a deeper understanding of genetic issues in drug response.
    B. It allows integration of different kinds of data with genomic data.
    C. It helps in personalizing treatments based on individual genetics.
    D. All of the above.

Answer: D) all of the above

Data science plays a crucial role in disease research. It not only provides a deeper understanding of genetic issues in drug response but also allows for the integration of different kinds of data with genomic data. This integration helps in personalizing treatments based on individual genetics. So, all the options are correct, making option D the most accurate answer.

14. Which search engine(s) make use of data science algorithms?
- A. Google only
- **B. Yahoo, Bing, Ask, AOL, and Google**
- C. Yahoo and Bing only
- D. AOL, Ask, and Google only

15. How does targeted advertising differ from traditional advertising?
- A. It uses data science algorithms to deliver personalized ads.
- B. It relies on user's past behavior to determine ad placement.
- C. It has a higher Call-Through Rate (CTR).
- **D. All of the above.**

Answer D) Targeted advertising is a method of placing ads based on demographics, consumers' previous buying history or preferences. It uses data science algorithms to analyze user behavior and deliver personalized ads, which results in a higher Call-Through Rate (CTR). So, all the options are correct, making option D the most accurate answer.

16. Which companies use website recommendations to improve user experience?
- A. Amazon, Twitter, and Google Play
- B. Netflix, LinkedIn, and IMDB
- **C. All of the above**
- D. None of the above

17. What can airline companies do with the help of data science?
- A. Predict flight delays
- B. Determine which class of airplanes to buy
- C. Decide whether to take a halt during a flight
- **D. All of the above**

18. From the given options, which statement accurately describes data science?
- A. Data science is a new field that solely focuses on analyzing data.
- B. Data science is used only in AI applications.
- **C. Data science helps in making machines intelligent by analyzing data.**
- D. Data science has no major applications in today's world.

19. How did data science help banking companies reduce losses?
- A. By collecting customer profiles and past expenditures
- B. By dividing and analyzing customer data variables
- C. By pushing banking products based on customer purchasing power
- **D. All of the above**

20. What is the main goal of genetics and genomics research?
- A. To understand the impact of DNA on our health only
- B. To acquire reliable personal genome data
- **C. To find individual biological connections between genetics, diseases, and drug response**
- D. All of the Above

21. What is one benefit of using data science in finance?
- A. **It helps in reducing bad debts and losses**.
- B. It allows companies to collect more data for future use.
- C. It improves customer service in banking industry.
- D. It enables companies to offer heavy discounts to customers.

22.**What is data collection?**
   A. A new concept introduced in our society
   B. A tedious process that requires technological knowledge
   C. An exercise that does not require technological knowledge
   D. An exercise that involves analyzing numbers and alpha-numerical data

23.**What does Data Science do?**
   A. Provides a clearer idea around the dataset
   B. Maintains records in institutions
   C. Incorporates AI into the data analysis process
   D. Generates predictions and suggestions by machines

24.**Which type of data is commonly used in data domain-based projects?**
   A. Textual data
   B. Graphical data
   C. Numerical or alpha-numerical data
   D. Audio-visual data

25.**What kind of databases are commonly found in banks?**
   A. Databases related to movie details
   B. Databases related to Salary of Teachers
   C. Databases related to employee registrations
   D. Databases related to locker owners

26.**What are the two ways in which data collection can be categorized?**
   A. Sensory and non-sensory collection methods
   B. Offline and online data collection methods
   C. Government and private sector collection methods
   D. Primary and secondary sources of data collection

27.**Which of the following is an example of an offline source of data collection mentioned in the text?**
   A. Sensors
   B. Open-sourced Government Portals
   C. Reliable Websites (Kaggle)
   D. World Organisations' open-sourced statistical Observations websites

28.**What should be kept in mind while accessing data from any data source?**
   A. Personal datasets can be used without the owner's consent
   B. Privacy breaches are acceptable for data collection
   C. Data should only be taken from reliable sources
   D. Data collected from random sources is always accurate

29.**What is the purpose of using reliable sources of data?**
   A. To ensure the authenticity of data for proper training of the AI model
   B. To collect personal datasets without consent
   C. To breach someone's privacy for data collection purposes
   D. To access confidential information for programming purposes

30.**In which format are tabular datasets commonly stored in Data Science?**
   A. CSV
   B. Spreadsheet
   C. SQL
   D. All of the above

31.**What does CSV stand for?**
   A. Comma separated values
   B. Computer spreadsheet values
   C. Centralized storage volumes
   D. Compressed structured variables

32. Which programming language is designed for managing data held in different kinds of DBMS?
    A. Python
    B. SQL
    C. Java
    D. C++

33. What is a Spreadsheet?
    A. A file format used to store tabular data
    B. A piece of paper or computer program used for recording and accounting data
    C. A programming language used for managing databases
    D. An online portal to access government statistics

34. Which Python package helps in accessing structured data (in tabular form)?
    A. CSV Reader
    B. Spreadsheet Parser
    C. SQL Integrator
    D. Pandas

35. Which of the following is NOT a guideline for data collection mentioned in the text?
    A. Use personal datasets without consent.
    B. Never breach someone's privacy to collect data.
    C. Take data from reliable sources.
    D. Use data available for public usage only.

36. What is the main focus of Data Science?
    A. Collecting data from various sources
    B. Analyzing and interpreting collected data
    C. Developing databases for record maintenance
    D. Just Implementing artificial intelligence in data analysis

37. How does AI contribute to the field of Data Science?
    A. It makes data collection easier and faster.
    B. It provides predictions and suggestions based on collected data.
    C. It eliminates the need for human involvement in analyzing data.
    D. It enhances technological knowledge for data collection.

38. Why should personal datasets only be used with the consent of the owner?
    A. To ensure accurate analysis and interpretation of the data
    B. To get better data when taken consent
    C. To maintain a fair balance between public and private information
    D. To prevent unauthorized access to personal information

39. NumPy, which stands for Numerical Python

40. _____is the fundamental package for Mathematical and logical operations on arrays in Python
    A. Pandas   **B. Numpy**   C. matplotlib   D none of the above

41. _____ is a multi-platform data visualization library built on NumPy arrays.
    A. Pandas   B. Numpy   **C. matplotlib**   D none of the above

2 Mark Questions

42. Ajay wants to access data from various sources. Suggest him any two points that he needs to keep in mind while accessing data from any data source.

Ans: While accessing data from any of the data sources, following points should be kept in mind:

1. Data which is available for public usage only should be taken up.

 2. Personal datasets should only be used with the consent of the owner.

3. One should never breach someone's privacy to collect data.

4. Data should only be taken from reliable sources as the data collected from random sources can be wrong or unusable.
5. Reliable sources of data ensure the authenticity of data which helps in the proper training of the AI model.
6. Data should be relevant to the problem. **(any two; 1 mark for each valid point)**

43. List down four online and offline sources of data from where we can collect data for AI projects

Offline Data Collection	Online Data Collection
Sensors	Open-sourced Government Portals
Surveys	Reliable Websites (Kaggle)
Interviews	World Organisations' open-sourced statistical websites
Observations	

44.List down any two formats I which tabular data sets can be stored.
The tabular datasets can be stored in different formats. Some of the commonly used formats are:
1. **CSV:** CSV stands for comma separated values. It is a simple file format used to store tabular data. Each line of this file is a data record and reach record consists of one or more fields which are separated by commas. Since the values of records are separated by a comma, hence they are known as CSV files.
2. **Spreadsheet:** A Spreadsheet is a piece of paper or a computer program which is used for accounting and recording data using rows and columns into which information can be entered. Microsoft excel is a program which helps in creating spreadsheets.
3. **SQL:** SQL is a programming language also known as Structured Query Language. It is a domain-specific language used in programming and is designed for managing data held in different kinds of DBMS (Database Management System) It is particularly useful in handling structured data.
(any two; 1 mark for each valid point)

45. Write notes on any one python packages which is used to access structured data(tabular format) inside the code.
1. NumPy: NumPy, which stands for Numerical Python, is the fundamental package for Mathematical and logical operations on arrays in Python. It is a commonly used package when it comes to working around numbers. NumPy gives a wide range of arithmetic operations around numbers giving us an easier approach in working with them. NumPy also works with arrays, which is nothing but a homogenous collection of Data.
2. **Pandas:** It is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. The name is derived from the term "panel data",

Pandas is well suited for many different kinds of data:
• Tabular data with heterogeneously-typed columns, as in an SQL table or Excel spreadsheet
• Ordered and unordered (not necessarily fixed-frequency) time series data.
• Arbitrary matrix data (homogeneously typed or heterogeneous) with row and column labels
• Any other form of observational / statistical data sets. The data actually need not be labelled at all to be placed into a Pandas data structure
3. Matplotlib: It is a visualization library in Python for 2D plots of arrays. Matplotlib is a multi-platform data visualization library built on NumPy arrays. One of the greatest benefits of visualization is that it allows us visual access to huge amounts of data in easily digestible visuals. Matplotlib comes with a wide variety of plots. Plots helps to understand trends, patterns, and to make correlations. They're typically instruments for reasoning about quantitative information.