S D M Institute of Technology (SDMIT), Ujire
DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

VI SEM   **Assignment Quiz (Objective Type Questions)**

Student Name:	USN:	Total Marks: 40
Course Name: Data Science and Visualization	Course Code: 21CS644	Month/Year: July/2024

Answer all the following questions. (Bubble the circle for a correct answer at the last page)

1. What is one of the reasons for the hype around Big Data and Data Science mentioned in the text?

a) Exaggerated claims about the importance of data scientists	b) Lack of recognition for previous work in related fields
c) Confusion with existing fields like statistics	d) All of the above

2. Which of the following is NOT one of the key factors that explains the significance of the current time period for the emergence of data science?

a) Availability of massive amounts of data	b) Abundance of inexpensive computing power
c) Advancements in quantum computing	d) Datafication of offline behaviors

3. What does "datafication" refer to?

a) The process of converting data into different formats	b) The process of quantifying and recording various activities and behaviors
c) The process of cleaning and preprocessing data	d) The process of visualizing data

4. According to the Venn diagram by Drew Conway, what are the three overlapping skill sets that define data science?

a) Statistics, Machine Learning, and Visualization	b) Hacking Skills, Math, and Business Knowledge
c) Statistics, Hacking Skills, and Domain Expertise	d) Coding, Math, and Visualization

5. What role can social scientists play in data science, according to the text?

a) Contributing expertise in understanding and analyzing human/user behavior data	b) Developing new statistical techniques for data analysis
c) Building data storage and processing infrastructure	d) Designing user interfaces for data products

6. Which of the following is NOT considered one of the key characteristics of Big Data?

a) Volume	b) Variety
c) Velocity	d) Veracity

7. According to Steve Lohr's definition, what does "Big Data" refer to as a "potential revolution"?

a) A revolution in measurement	b) A revolution in computing power
c) A revolution in data storage	d) A revolution in data visualization

8. What is the purpose of statistical inference, according to the text?

a) To collect and store large amounts of data	b) To understand and make sense of complex, random, and uncertain real-world processes
c) To develop new mathematical models	d) To visualize data in meaningful ways

9. What is the difference between a population and a sample in statistics?

a) A population is a subset of a sample	b) A sample is a subset of a population
c) A population and a sample are the same thing	d) There is no difference between a population and a sample

10. Which assumption or pitfall related to Big Data analysis is mentioned in the text?

a) The assumption that Big Data is always accurate and error-free	b) The assumption that Big Data can only be analyzed using specialized software

	c) The claim that with Big Data we have "N=ALL" or all the data	d) The assumption that Big Data is only useful for large corporations
11	**What is the first step in the data science process?**	
	a) Data Cleaning	b) Data Collection
	c) Data Modeling	d) Data Visualization
12	**What is the main goal of feature selection in predictive models?**	
	a) Increase the number of features	b) Improve model interpretability
	c) Increase computational cost	d) Complicate the model
13	**Which method evaluates the relevance of each feature independently of the learning algorithm?**	
	a) Wrapper Methods	b) Filter Methods
	c) Embedded Methods	d) Recursive Feature Elimination
14	**What is a common example of a wrapper method for feature selection?**	
	a) LASSO regression	b) Pearson correlation
	c) Recursive Feature Elimination (RFE)	d) Principal Component Analysis (PCA)
15	**What is a key challenge in building recommendation systems at scale?**	
	a) Small data handling	b) Linear algebra knowledge
	c) Simple coding	d) Minimal data usage
16	**What was the initial business model of Hunch.com?**	
	a) API model	b) Personalized advice through a questionnaire
	c) Social networking	d) E-commerce platform
17	**What major company acquired Hunch?**	
	a) Amazon	b) Google
	c) Facebook	d) eBay
18	**Which technique is used to simplify complex data by reducing the number of dimensions while preserving important information?**	
	a) Linear Regression	b) Feature Selection
	c) Principal Component Analysis (PCA)	d) Naive Bayes
19	**What is the goal of PCA in data analysis?**	
	a) Increase data dimensions	b) Preserve data complexity
	c) Reduce data dimensions while preserving variance	d) Add noise to the data
20	**What is Singular Value Decomposition used for in data analysis?**	
	a) To increase data dimensions	b) To identify patterns in the data
	c) To eliminate features	d) To increase data complexity
21	**Which of the following is not a step in SVD?**	
	a) Decompose the data matrix	b) Identify singular values
	c) Reconstruct the original data	d) Add noise to the data
22	**What is a bipartite graph used for in recommendation systems?**	
	a) To connect users to users	b) To connect items to items
	c) To connect users to items	d) To disconnect users from items
23	**Which algorithm is often used to improve recommendations over time by analyzing user responses?**	

	a) Decision Trees	b) k-Nearest Neighbors
	c) Machine Learning algorithms	d) Random Forest
24	**How does PCA help in improving model performance?**	
	a) By adding more features	b) By increasing data redundancy
	c) By reducing noise and simplifying the data	d) By complicating the model
25	**What is the main purpose of data visualization?**	
	a) To store data	b) To make data understandable
	c) To collect data	d) To delete data
26	**Which Python library is primarily used for data analysis?**	
	a) NumPy	b) pandas
	c) Matplotlib	d) seaborn
27	**What is the first step in the data wrangling process?**	
	a) Analyzing data	b) Cleaning data
	c) Importing data as a DataFrame	d) Transforming data into graphs
28	**Which measure of central tendency is calculated by summing all measurements and dividing by the number of observations?**	
	a) Mean	b) Median
	c) Mode	d) Range
29	**What type of data visualization shows the relationship between body mass and maximum longevity of various animals?**	
	a) Line chart	b) Bar chart
	c) Scatter plot	d) Pie chart
30	**Which of the following is a measure of dispersion?**	
	a) Mean	b) Median
	c) Mode	d) Standard deviation
31	**In a positive correlation, how do the variables move?**	
	a) In opposite directions	b) Independently
	c) In the same direction	d) Randomly
32	**What type of chart would be best to visualize the frequency distribution of a single variable?**	
	a) Pie chart	b) Histogram
	c) Line chart	d) Scatter plot
33	**What does the term "data mining" refer to?**	
	a) The process of cleaning data	b) The exploration and analysis of large datasets to discover patterns
	c) The visualization of data	d) The storage of data
34	**Which Python library provides a high-level interface for drawing attractive and informative statistical graphics?**	
	a) pandas	b) Matplotlib
	c) seaborn	d) NumPy
35	**What is the top-level container for all plot elements in Matplotlib?**	
	a) Axes	b) Figure
	c) Subplot	d) Canvas

36	Who developed Matplotlib?	
	a) Guido van Rossum	b) John Hunter
	c) Bjarne Stroustrup	d) James Gosling
37	Matplotlib was initially created to emulate which software?	
	a) Excel	b) R
	c) SPSS	d) MATLAB
38	What Python object contains multiple plots within a single visualization?	
	a) Axes	b) Figure
	c) Subplot	d) Canvas
39	Which component of a Matplotlib plot contains the actual plot and its elements?	
	a) Figure	b) Axes
	c) Spine	d) Grid
40	What is used to describe the content of the plot in Matplotlib?	
	a) Spine	b) Title
	c) Legend	d) Grid

1.	Ⓐ Ⓑ Ⓒ Ⓓ	21.	Ⓐ Ⓑ Ⓒ Ⓓ
2.	Ⓐ Ⓑ Ⓒ Ⓓ	22.	Ⓐ Ⓑ Ⓒ Ⓓ
3.	Ⓐ Ⓑ Ⓒ Ⓓ	23.	Ⓐ Ⓑ Ⓒ Ⓓ
4.	Ⓐ Ⓑ Ⓒ Ⓓ	24.	Ⓐ Ⓑ Ⓒ Ⓓ
5.	Ⓐ Ⓑ Ⓒ Ⓓ	25.	Ⓐ Ⓑ Ⓒ Ⓓ
6.	Ⓐ Ⓑ Ⓒ Ⓓ	26.	Ⓐ Ⓑ Ⓒ Ⓓ
7.	Ⓐ Ⓑ Ⓒ Ⓓ	27.	Ⓐ Ⓑ Ⓒ Ⓓ
8.	Ⓐ Ⓑ Ⓒ Ⓓ	28.	Ⓐ Ⓑ Ⓒ Ⓓ
9.	Ⓐ Ⓑ Ⓒ Ⓓ	29.	Ⓐ Ⓑ Ⓒ Ⓓ
10.	Ⓐ Ⓑ Ⓒ Ⓓ	30.	Ⓐ Ⓑ Ⓒ Ⓓ
11.	Ⓐ Ⓑ Ⓒ Ⓓ	31.	Ⓐ Ⓑ Ⓒ Ⓓ
12.	Ⓐ Ⓑ Ⓒ Ⓓ	32.	Ⓐ Ⓑ Ⓒ Ⓓ
13.	Ⓐ Ⓑ Ⓒ Ⓓ	33.	Ⓐ Ⓑ Ⓒ Ⓓ
14.	Ⓐ Ⓑ Ⓒ Ⓓ	34.	Ⓐ Ⓑ Ⓒ Ⓓ
15.	Ⓐ Ⓑ Ⓒ Ⓓ	35.	Ⓐ Ⓑ Ⓒ Ⓓ
16.	Ⓐ Ⓑ Ⓒ Ⓓ	36.	Ⓐ Ⓑ Ⓒ Ⓓ
17.	Ⓐ Ⓑ Ⓒ Ⓓ	37.	Ⓐ Ⓑ Ⓒ Ⓓ
18.	Ⓐ Ⓑ Ⓒ Ⓓ	38.	Ⓐ Ⓑ Ⓒ Ⓓ
19.	Ⓐ Ⓑ Ⓒ Ⓓ	39.	Ⓐ Ⓑ Ⓒ Ⓓ
20.	Ⓐ Ⓑ Ⓒ Ⓓ	40.	Ⓐ Ⓑ Ⓒ Ⓓ

Total Marks Obtained Out of 40:	Total Marks Obtained Out of 20:
Student Signature:	Faculty Name and Signature: