Reg. No.															

Question Paper Code	13711

B.E. / B.Tech. - DEGREE EXAMINATIONS, APRIL / MAY 2025

Third Semester

Artificial Intelligence and Data Science

20AIPW301 – FUNDAMENTAL OF DATA SCIENCE WITH LABORATORY

Regulations - 2020

Duration: 3 Hours | Max. Marks: 100

PART - A (MCQ) (10 × 1 = 10 Marks)

Answer ALL Questions

		Marks	K – Level	CO
1.	What is the primary purpose of data collection in data science? (a) To store data for future use (b) To ensure data privacy **(**c) To gather relevant information that will be analyzed to gain insights (d) To clean the data	1	K1	CO1
2.	Which level of data provides actionable insights that can be used to make decisions? (a) Raw Data (b) Processed Data (c) Information (d) Knowledge	1	K1	CO1
3.	Which feature in Excel allows you to import data from various external sources? (a) Data Validation (b) Get & Transform Data (c) Conditional Formatting (d) Pivot Tables	1	K1	CO2
4.	What is a dashboard in the context of spreadsheets? (a) A tool for data entry (b) A visual display of key information (c) A feature for importing data (d) A method for data validation	1	K1	CO2
5.	Which of the following is a typical step in data preparation? (a) Data mining (b) Data imputation (c) Data modelling (d) Data encryption	1	K1	CO3
6.	Which library is commonly used in Python for numerical operations involving matrices and vectors? (a) Pandas (b) NumPy (c) NetworkX (d) SciPy	1	K1	CO3
7.	Which of the following tools is widely used for data munging and manipulation? (a) Jupyter Notebook (b) SPSS (c) Python (Pandas library) (d) Google Sheets	1	K1	CO4
8.	Which data visualization method is best for identifying the outliers in a dataset? (a) Histogram (b) Scatter Plot (c) Box Plot (d) Bar Chart	1	K1	CO4
9.	Which phase in the Big Data Life Cycle involves converting raw data into structured formats? (a) Acquisition (b) Storage (c) Analysis (d) Visualization	1	K1	CO5
10.	Which type of model is often used for predicting stock prices in finance? (a) Linear regression (b) Decision Trees (c) Time series analysis (d) Clustering	1	K1	CO6

PART - B (12 × 2 = 24 Marks)

Answer ALL Questions

11.	What is data science?	2	K1	CO1
12.	State the 3V model in big data.	2	K1	CO1
13.	Define spreadsheet.	2	K1	CO2
14.	Mention the VLOOKUP different from the feature of lookup.	2	K1	CO2
15.	List any five methods in pandas.	2	K1	CO3
16.	How to produce scatter plot in python?	2	K1	CO3
17.	Why is data visualization important?	2	K1	CO4
18.	Mention two best features of any two data science tools.	2	K1	CO4

19.		Interpret the SMAM model.	*2*	*K2*	*CO5*
20.		Explain the ethical issues in data science?	*2*	*K2*	*CO5*
21.		Write short notes on bioinformatics used in machine learning.	*2*	*K1*	*CO6*
22.		What is the application of data science in process optimization?	*2*	*K1*	*CO6*

PART - C (6 × 11 = 66 Marks)

Answer ALL Questions

23.	a)	Discuss the quantitative and qualitative data with an example.	*11*	*K2*	*CO1*
		OR			
	b)	Elaborate the relationship between data warehouse and data science.	*11*	*K2*	*CO1*
24.	a)	Demonstrate the importing data into spreadsheet from different data sources with an example.	*11*	*K3*	*CO2*
		OR			
	b)	Illustrate the data visualization in spread sheet with an example.	*11*	*K3*	*CO2*
25.	a)	Explain in detail about the Principal Component Analysis (PCA) in detail with an example.	*11*	*K2*	*CO3*
		OR			
	b)	Explain the process of data cleaning and preliminary data analysis.	*11*	*K2*	*CO3*
26.	a)	Interpret in detail about the data visualization tools with an example.	*11*	*K2*	*CO4*
		OR			
	b)	Explain in detail about the data mungling and modelling tools in data science.	*11*	*K2*	*CO4*
27.	a)	Discuss about the sequential steps in SEMMA methodology.	*11*	*K2*	*CO5*
		OR			
	b)	Describe the big data life cycle with neat diagram.	*11*	*K2*	*CO5*
28.	a)	Demonstrate the data preprocessing for the employee data set in python.	*11*	*K3*	*CO6*
		OR			
	b)	Illustrate the various applications used in data science fields.	*11*	*K3*	*CO6*