Reg. No.												

Question Paper Code	11694

B.E./B.Tech. - DEGREE EXAMINATIONS, NOV/DEC 2022

Third Semester

Artificial Intelligence and Data Science

20AIPW301 -- FUNDAMENTAL OF DATA SCIENCE WITH LABORATORY

(Regulations 2020)

Duration: 3 Hours

Max. Marks: 100

PART - A (10 × 2 = 20 Marks)

Answer ALL Questions

		Marks, K-Level,CO
1.	Define data science.	2,K1,CO1
2.	List out the areas in which Data Science can be applied.	2,K1,CO1
3.	What are the aggregation functions available in NumPy?	2,K1,CO2
4.	How will you save the figure to a file in python?	2,K1,CO2
5.	What is Principal Component Analysis (PCA) and How it is used?	2,K1,CO3
6.	Define data quality.	2,K1,CO3
7.	Why data visualization is important? List out the tools used for data visualization?	2,K1,CO4
8.	List the benefits of data modelling.	2,K1,CO4
9.	What are the three major ethical principles for data scientists?	2,K1,CO5
10.	What are the big challenges of big data?	2,K1,CO5

PART - B (5 × 13 = 65 Marks)

Answer ALL Questions

11.	a)	Explain in detail about the data preprocessing with an example.	13,K2,CO1
		OR	
	b)	Discuss the following with examples (i) Structured, Unstructured and Semi-Structured. (ii) Quantitative and Qualitative data. (iii) The four Levels of Data.	5,K2,CO1 4,K2,CO1 4,K2,CO1
12.	a)	Explain in detail about any four python libraries that are used in Data Science.	13,K2,CO2
		OR	
	b)	Demonstrate the importing data into spreadsheet from different data sources with an example.	13,K3,CO2

13. a) Explain the process of data cleaning and preliminary data analysis. *13,K2,CO3*

OR

b) What is the need of dimensionality reduction? Explain PCA techniques for dimensionality reduction with examples. Consider a set of 2D points {(-3,-3), (-1,-1),(1,1),(3,3)}. We want to reduce the dimensionality of these points by 1 using PCA algorithms. Assume sqrt(2)=1.414. *13,K3,CO3*

14. a) Describe in detail about the components of spread sheets. *13,K2,CO4*

OR

b) Prepare a dashboard for IPL dataset using excel. *13,K3,CO4*

15. a) Explain in detail about the different phase in CRISP-DM Methodology. *13,K2,CO5*

OR

b) Describe the Big data life cycle with neat diagram. *13,K2,CO5*

PART C (1 × 15 = 15 Marks)

16. a) Explain the various applications used in data science. *15,K2,CO6*

OR

b) Analyze the data model for student performance analysis using python. *15,K4,CO6*