Question Paper Code

13122

B.E. / B.Tech. - DEGREE EXAMINATIONS, NOV / DEC 2024

Third Semester

Artificial Intelligence and Data Science

20AIPW301 - FUNDAMENTAL OF DATA SCIENCE WITH LABORATORY

Regulations - 2020

Dυ	ration: 3 Hours	Iax. Mar	ks: 1	00
	PART - A (MCQ) $(20 \times 1 = 20 \text{ Marks})$			
	Answer ALL Questions	Marks	Level	co
1.	Which of the following is the most important language for Data Science?	1	<i>K1</i>	CO1
	(a) Java (b) Ruby (c) R (d) None of the mentioned			
2.	Which type of data is best suited for statistical analysis and creating numerical models?	1	K1	CO1
	(a) Qualitative data (b) Quantitative data			
	(c) Both qualitative and quantitative data equally (d) Neither			
3.	Which of the following is a common concern in data science?	1	<i>K1</i>	CO1
	(a) Data privacy and security			
	(b) Colour theory in data visualization			
	(c) Hardware requirements for data storage			
	(d) Programming languages used in database management			~~-
4.	What does a correlation coefficient of 0 indicate?	1	KI	CO2
	(a) No correlation (b) Perfect positive correlation			
_	(c) Perfect negative correlation (d) Strong positive correlation	0 1	1/1	G02
5.	Which Excel function is used to look up a value in a table based on a specific condition	? 1	<i>K1</i>	CO2
_	(a) VLOOKUP (b) HLOOKUP (c) INDEX (d) MATCH	1	K1	CO2
6.	Which Excel feature allows you to summarize data with dynamic tables?		ΚI	CO2
7	(a) PivotTables (b) Charts (c) Conditional Formatting (d) Data Validati	on 1	<i>K1</i>	CO3
7.	What is the primary purpose of Principal Component Analysis (PCA)? (a) To perform clustering on high-dimensional data	1	IX I	COS
	(b) To reduce the dimensionality of data while retaining as much variance as possible			
	(c) To predict future values based on historical data			
	(d) To create a data visualization dashboard			
8.	In data science, which of the following is a common representation for features of data	in 1	<i>K1</i>	CO3
	machine learning models?			
	(a) Scalars (b) Vectors (c) Strings (d) Graphs			
9.	What is a common challenge associated with high-dimensional data?	1	<i>K1</i>	CO3
	(a) Lack of data sources (b) Increased computational cost			
	(c) Data collection difficulty (d) Limited storage capacity			
10.	Which of the following tools is primarily used for machine learning and modeling in Da	ata 1	K1	CO4
	Science?			
	(a) Tensor Flow (b) Microsoft Word (c) Canva (d) Power BI			
11.	Which type of visualization is best used for showing the distribution of a dataset?	1	K1	CO4
	(a) Bar Chart (b) Box Plot (c) Pie Chart (d) Line Graph			~~.
12.	Which of the following is an advanced data visualization tool?	1	<i>K1</i>	CO4
10	(a) Excel (b) Tableau (c) Notepad (d) Word Press	,	1/1	G05
13.	What does the acronym SMAM stand for in the context of data science?	1	K1	CO5
	(a) Structured Modeling and Management (b) Sample Management and Model			
1.4	(c) Sample, Modify, Analyze, Manage (d) Survey, Model, Apply, Measure	1	<i>K1</i>	CO5
14.	Which of the following is a concern for data scientists when working with user data? (a) High-quality data collection (b) Ethical concerns regarding data misus		ΚI	003
	(c) Increasing hardware capacity (d) Automating data workflows)C		
	(u) Automating data workflows			
1/1			12	122

15.		ch of the following is a risk of poor data handling by data scientists? Overfitting of models (b) Violation of privacy and data breach	1	K1	CO5
16.	Whi	improved business outcomes (d) Efficient data processing ch phase of CRISP-DM involves understanding the business objectives and straints?	1	K1	CO5
		Data Preparation (b) Business Understanding			
		Evaluation (d) Deployment			
17.	` ′	ch of the following can data analytics in education help with?	1	<i>K1</i>	CO6
		Discovering new planets (b) Predicting student dropout rates			
		Automating financial transactions (d) Designing engineering products			
18		ch technology is often combined with data science for smart city infrastructure	1	<i>K1</i>	CO6
10.		agement?			
		nternet of Things (IoT) (b) Genetic algorithms			
		Blockchain (d) Quantum computing			
10	` /	ch of the following data science techniques is often used for real-time player	1	<i>K1</i>	CO6
19.		chmaking in online games?	•	111	
		Clustering (b) Reinforcement learning			
	(c) I	Regression analysis (d) Data annotation			
20.		ch data science technique is often used to analyze student performance data in	1	<i>K1</i>	<i>CO6</i>
		cation?			
	(a) T	Time series analysis (b) Regression analysis			
	(c) 1	Natural Language Processing (NLP) (d) Reinforcement learning			
		$PART - B (10 \times 2 = 20 \text{ Marks})$			
		Answer ALL Questions			
21.	Wha	at is the difference between Big Data and Data Science?	2	<i>K1</i>	CO1
		ine Data Transformation.	2	<i>K1</i>	CO1
		y is it so important to pre-process the data before loading it?	2	<i>K1</i>	CO2
		at is the difference between opening an Excel file and importing it?	2	<i>K1</i>	CO2
		at is a data frame give example?	2	<i>K1</i>	CO3
		at is Data Acquisition?	2	<i>K1</i>	CO3
		ine data wrangling.	2	K1	
		y data visualization is important?	2	K1	
	•	•	2	K1	CO5
	. What is SEMMA methodology?		2		CO6
30.	Hov	v is machine learning used in bioinformatics?	2	KI	COO
		PART - C $(6 \times 10 = 60 \text{ Marks})$			
21	۵)	Answer ALL Questions Define data seignes, Discuss the amplications of data seignes with examples	10	K2	CO1
31.	a)	Define data science. Discuss the applications of data science with examples.	10	IX2	COI
	1. \	OR	10	K2	CO1
	b)	Explain the Relationship between data warehouse and data science.	10	K2	COI
22	,		10	V2	cor
32.	a)	Demonstrate the Pivot Tables and Charts with an exampleVLOOKUP-Dashboard	10	K2	CO2
		in Spread sheet.			
	• .	OR	10	77.0	G02
	b)	Illustrate the Data Cleaning and Preliminary Data Analysis with an example.	10	<i>K2</i>	CO2
			10	****	a
33.	a)	Explain in detail about data acquisition and its process.	10	K2	CO3
		OR	10	***	~~
	b)	Explain Data Modeling Process.	10	K2	CO3
_					~~:
34.	a)	Describe about plotting and visualization concepts in python with suitable code and	10	K2	CO4
		examples.			
		OR			

OR

	b)	Explain in details about	10	K2	CO4
		(i) Elementary visualization means			
		(ii) Box plots			
		(iii) Charts			
		(iv) Graphs			
35.	a)	Explain the various methodologies used in data analytics.	10	K2	CO5
		OR			
	b)	Discuss about the Ethical guidelines for Data Scientist.	10	K2	CO5
36.	a)	Apply the function in Excel to find the saving percentage of the person who spent 10	10	К3	CO6
		% of his salary for rent, 20% of remaining part of his salary for food after which he spent 50 % balance of the salary on other expenditure.			
		OR			
			10	77.3	and
	b)	Apply a data analytics for employee database data set.	10	K3	CO6