



NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY
(AN AUTONOMOUS INSTITUTE AFFILIATED TO VTU, BELAGAVI)

July / August 2024 Sixth Semester End Examination BE Degree
Department of Computer Science and Engineering
Data Science using Python (21CSE652)

Max. Marks:100

Duration: 3 Hrs

Instructions

1. Part A and Part C – Answer all questions
2. Part B - Answer one full question from each unit
3. Missing Data (if any) can be suitably assumed

PART A			Marks	CO; BL
1.	a.	Differentiate between list and tuple in Python	3	1;2
	b.	Define data cleaning. List any four data cleaning operations.	3	2;1
	c.	List the salient features of box plot.	3	3;1
	d.	Differentiate between supervised and unsupervised learning.	3	4;2
	e.	List the different activation functions used in deep neural networks.	3	5;1
PART B				
UNIT I				
2.	a.	Consider a weather monitoring system that collects temperature readings across Bangalore. Temperature data is stored as 2D Numpy array, where each row represents temperatures for a specific day. Calculate average temperature recorded by each sensor for a week, applying the python numerical computation methods.	7	1;3
	b.	Bring out the significance of pandas. Mention any three data manipulation operations that can be done using pandas. Give suitable examples for the same.	7	1;2
OR				
3.	a.	When developing a student performance record, it is observed that a python dictionary could have course code as key and performance score as value for a single student. For one student, how do you find average score of the student across several subjects? Construct an example.	7	1;3
	b.	Provide summary of the following: i) Python sets ii) Python functions Give suitable examples for the same.	7	1;2
UNIT II				
4.	a.	Consider a DataFrame in Titanic dataset. The dataset has PassengerId, survived, pclass, name, sex, age, sibsp, parch, ticket, fare, cabin and embarked columns. It is observed that age and cabin columns contain missing fields. Using pandas create a dataframe and handle missing data of age and cabin in this dataset.	7	2;3
	b.	"Parsing of JSON dataset using pandas is very convenient." Comment on this statement with suitable examples.	7	2;2
OR				
5.	a.	Consider a scenario where your college is planning to organize Alumni meet for 2010 B.E passed out students. The contact details and addresses of these alumni are available in a MySQL database named Alumni2010. Write a python script to connect to the database and retrieve their contact details and addresses.	7	2;3
	b.	Bring out the significance of slicing and indexing concepts in python, Give suitable examples.	7	2;2
UNIT III				
6.	a.	Consider two python lists. The first list contains a list of movies and the second list contains number of Oscars received by the corresponding movie in the first list. Using matplotlib, explain how bar charts and line charts are drawn.	7	3;2
	b.	Develop an interactive Bokeh plot that shows the sales performance of different products over time.	7	3;2

OR

7. a. Is it possible to build interactive charts and plots with python? Give reasons for your answer and elaborate with an example. 7 3;2
- b. Imagine three lists. 'x' list has values in the range 1 to 10. 'y' list has values that are 10 times that of each item of x list. 'z' list has values that are sum of corresponding values of 'x' and 'y' list. With these three lists, demonstrate how 3D plots can be generated using Matplotlib. 7 3;2

UNIT IV

8. a. Enumerate the steps in the working of KNN algorithm. State its advantages and disadvantages. Give a suitable example. 8 4;2
- b. "SVMs are fast and more accurate than logistic regression". Substantiate above statement with suitable examples. 6 4;2

OR

9. a. "Regression is a form of predictive modeling which investigates the relationship between dependent and independent variable." In light of this given statement, elaborate and compare linear, polynomial and logistic regression. 8 4;2
- b. Elaborate on decision tree algorithm with a simple example. 6 4;2

UNIT V

10. a. Enumerate the design steps of recurrent neural networks for vector input sequences conditional on the context with an example. 8 5;2
- b. Provide an overview of deep neural network with a neat diagram. 6 5;2

OR

11. a. Elaborate on Generative Adversarial Network (GAN). What is its significance? 8 5;2
- b. Compare CNN and RNN; List their applications. 6 5;2

PART C

12. Consider the tabulated data given below. 5 4;4

Height	Age	Weight
5	45	59
5.11	46	77
5.6	55	61
5.7	56	65
5.4	28	50
5.3	35	53
5.0	46	59

Now, with the value of $k=3$, elaborate a detailed step by step solution to determine weight for this data, height =5.5 and age=38 using KNN algorithm

13. In a financial forecasting scenario, consider a dataset with columns such as dates revenue and expenses. Apply appropriate pandas library methods to summarize revenues and expenses. Also, analyze monthly cash flow trends. 5 1;4
14. Analyze the architectural structure of GAN and substantiate how GANs can produce generative outputs with a neat diagram. 5 5;4