21CSE652 USN

NITTE EDUCATION TRUST

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)

Duration: 3 Hrs **Max. Marks:100**

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 lists and tuples in python. Give examples for each.	3	1;2
	b.	Illustrate slicing and indexing method in the context of data processing with suitable examples.	3	2;2
	c.	Provide an overview of Line chart, Barplot and Boxplot in python.	3	3;2
	d.	Summarize about linear regression; List its applications and advantages.	3	4;2
	e.	Outline the concept of Deep learning; paraphrase its scope and applications.	3	5;2
		PART B		
		UNIT I		
2.	a.	Identify various data types in python, and also provide variable naming conventions	8	1;3
	b.	Explicate the different operators for arithmetic operations and relational operations in Python, with illustrative examples.	6	1;2
		OR		
3.	a.	Consider a scenario in which you need to find out 'first class' students, each scoring 60% and above marks in a subject. Develop python code for this purpose and identify the control structures used in this.	8	1;3
	b.	How can we do data exploration-using pandas? Consider panda's Data Frame and Series and explain the concepts.	6	1;2
		UNIT II		
4.	a.	Which are the steps used for data cleaning in pandas. Illustrate the process.	7	2;2
	b.	Usually big data sets are stored in CSV or JSON format. With specific examples, show how to read CSV and JSON files having data, using pandas library. Show program steps to read the data and convert into Data Frame format from .CSV/.JSON	7	2;2
		OR		
5.	a.	Interpret and illustrate the working of the following pandas functions on Data Frames. i. read ii. shape iii. dropna iv. describe v. concat vi. merge vii. sum	7	2;2
	b.	Using Pandas how can we work with missing data? Elaborate and give examples.	7	2;2
		UNIT III		
6.	a.	Construct boxplot and scatterplot using matplotlib, taking example data of student's scores in a particular test. Derive the various useful interpretations from these plots.	8	3;3
	b.	A teacher covers certain portion of syllabus every day. Develop a time series plot with pandas to visualize the progress of the teaching date-wise. Explain the steps involved.	6	3;3

OR

7.	a.	Identify the method of constructing three dimensional plot using matplotlib. Mention the significance of these kinds of plots.	8	3;3
	b.	Apply seaborn for data visualization in python programs and illustrate the working with appropriate data values.	6	3;3

UNIT IV

8.	a.	Provide a summary of polynomial regression and logistic regression; put forth specific instances wherein these methods are applicable.	8	4;2
	b.	Model a Naïve Bayes classifier for predicting whether it is suitable to play cricket under some conditions of rainy, sunny, cloudy, day time or night time etc. Identify the advantages of a Naïve Bayes approach.	6	4;3
		OR		
9.	a.	Provide an overview of decision tree and random forest methods.	8	4;2
	b.	Identify the working method, and applications of Support Vector Machine algorithm.	6	4;3

UNIT V

10.	a.	With a neat diagram illustrate a fully connected feed forward deep neural network; elaborate on its working and about the component parts.	8	5;2
	b.	Why is recurrent neural network useful? Show its working with an example.	6	5;2
		OR		
11.	a.	Introduce Convolutional Neural Network (CNN) and elaborate on the architectural features of a typical CNN with a neat diagram.	8	5;2
	b.	Outline some of the advantages and disadvantages of deep learning neural networks.	6	5;2

PART C

12.		Analyze how we can do numerical computing using numpy. Write a program in Python using Numpy to evaluate a given expression comprising addition, multiplication, subtraction and division.	5	1;4
13.		When we need to predict if a student passes or fails in an exam given the number of hours of preparation. The target variable has two values, pass or fail. Analyze this problem and solve this using logistic regression to predict pass or fail classification.	5	4;4
14.		Compare and contrast convolutional neural network and recurrent neural network. Access their suitability for different applications, like computer vision , weather forecasting, language processing etc.	5	5;4