22CSE554 USN | | | | | | | | | | |

NITTE EDUCATION TRUST

NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY

(AN AUTONOMOUS INSTITUTE AFFILIATED TO VTU, BELAGAVI)

Dec .24 / Jan. 2025 Fifth Semester End Examination BE Degree

Department of Computer Science and Engineering

Data Science using Python (22CSE554)

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

			Marks	CO; BL
		PART A		
1.	a.	Explain how Numpy handles numerical operations on arrays more efficiently than python list with the suitable example.	3	1;2
	b.	Show two common techniques used in data cleaning to handle missing values in a data frame with the suitable example for each.	3	2;2
	c.	Illustrate a plot that is suitable to depict the relationship between two variables.	3	3;2
	d.	Explain the difference between supervised and unsupervised learning with an example for each.	3	4;2
	e.	Outline the role of activation functions in neural networks and why non-linear activation functions are typically used.	3	5;2
		PART B		
		UNIT I		
2.	a.	Write a python program to analyze a company's sales data. The program should take monthly sales data for a year as input, calculate total annual sales, the average monthly sales and identify the month(s) with the highest sales.	8	1;3
	b.	Using np.array() and np.where(), create a 2D array from a list of lists. Find the sum of elements in each row that are greater than 10.	6	1;2=3
		OR		
3.	a.	Identify the python libraries suitable for data analysis by highlighting at least 3 of its functions from each library.	8	1;3
	b.	Create an array with 15 elements. Reshape it into a 3X5 matrix and compute the median of each row.	6	1;3
		UNIT II		
4.	a.	Outline ay 4 read function in pandas with suitable example for each.	8	2;2
	b.	Create a pandas dataframe using MultiIndex with levels ['state 1; 'state 2'] and years [2000, 2010]. Assign population data to the MultiIndex and unstack it to transform the MultiIndex into columns.	6	2;3
		OR		
5.	a.	Show slicing and dicing operations in panda's dataframe with the suitable examples.	8	2;2
	b.	Using pd.concat(), concatenate two pandas series with different indices and display the resulting series. Set the axis parameter to perform column-wise concatenation.	6	2;3
		UNIT III		
6.	a.	Outline the purpose of pairplot, Heatmap, catplot and kdeplot in seaborn library with a suitable example for each.	8	3;2
	b.	Write a python program to create a line chart in Matplotlib to show the daily temperature variations over a week. Use different line style for weekdays and weekends. Add a legend to the plot.	6	3;3
		OR		
7.	a.	Using the 'iris' dataset, create an interactive funnel area chart with plotly to visualize the relationship between 'sepal-length' and 'species'. Set the following: Hover text, Title, and Template parameters.	8	3;2
	b.	Using Basemap in python, create a map centered on Karnataka, India. Set the projection to Lambert Conformal ('lcc'), with a center at latitude 15.3173 and longitude 75.7139, and a width and height of 2 million meters. Add a topological layer with a scale of 0.5 and alpha transparency of 0.5. Plot a marker for Karnataka's location and label it "Karnataka'. Display the map with an appropriate title. Write the code to implement this map.	6	3;3

UNIT IV

8. a. Consider the dataset with 'Age' and 'Income level' as predictors and 'purchase decision' as target variable. 7 4;3

	Age	Income level	Purchase decision
1	Young	Medium	Yes
2	Middle aged	High	Yes
3	Senior	Medium	No
4	Young	Low	No
5	Middle aged	Medium	Yes
6	Senior	High	No
7	Young	Medium	Yes
8	Senior	Medium	Yes
9	Middle aged	Low	No
10	young	High	Yes

Using Naïve Bayes classifier, classify whether 'young' customer with 'medium' income is likely to purchase or not.

b. Compare logistic regression and Naïve Bayes in terms of their approach to classification. Identify scenarios where one might be preferred over other. 7 4;2

OR

9. a. A real estate agency wants to predict the price of houses based on their size. The relationship between price Y (in Thousands of dollars) and size X= (in sq.ft) is given by; 7 4;3

Y=0.2X+50

i. Calculate the predicted price of a 1,200 sq.ft house.

ii. A 1,500 sq.ft house sold for $38,00,000(dollars) find the residual(error) between the predicted and actual price.

iii. By how much the price increases if the house size increases by 100 sq,ft. according to this model.

b. Outline the working of the KNN algorithm and its suitability for different data types. Explain one advantage and one disadvantage of using KNN in classification problem. 7 4;2

UNIT V

10. a. Consider a perceptron model with two inputs, x_1 and x_2 and the following parameters; 8 5;3

1. Weights : $W_1=1, W_2=1$
2. Bias: b= -1.5
3. Activation function: Step function, which output 1 if input is>=0 otherwise, 0;

and answer the following questions.

i. Determine the output of the perceptron for each of the following input pairs (x_1,x_2): (0,0), (0,1), (1,0), (1,1)

ii. Based on the results, identify the logical operation the perceptron is simulating.

b. Outline the linear separability problem. Show at least a solution to resolve the linear separability problem. 6 5;2

OR

11. a. Consider this simple ANN with one hidden layer 8 5;3

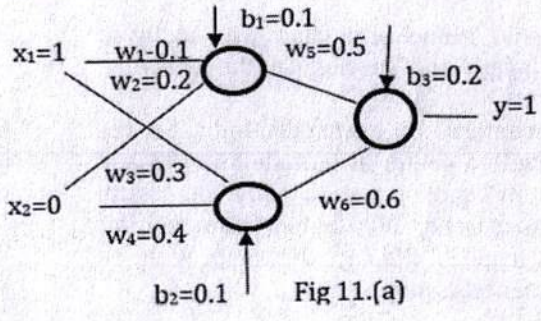

Fig 11.(a)

And answer the following question

i. Calculate the output of hidden layer neurons L1 and L2 using ReLU function.

ii. Calculate the final output of the network.

iii. Calculate the error using MSE.

b. Compare CNN and RNN in terms of their architecture and primary applications. Identify scenarios where one architecture would be preferred over the other. 6 5;3

PART C

12. Using Bokel, analyze and visualize a dataset of daily temperature readings over a month. Create a simple line chart that shows the trend in temperature across the days. 5 3;4

i. Add appropriate axis labels to represent the days and temperature.

ii. Ensure the plot can be displayed interactively in a jupyter notebook or web browser.

13. Evaluate the difference between Decision trees and Random forests. Explain how Random Forests address the limitations of Decision trees in machine learning applications. 5 4;4

14. Analyze the perceptron model and explain how it can be used to solve binary classification tasks such as AND and OR. Discuss the limitations of the perceptron in solving non-linear problems. 5 5;4