

**VIT**Vellore Institute of Technology
(Established as a University under section 3 of the UGC Act, 1956)**SCHOOL OF COMPUTER SCIENCE AND ENGINEERING**

Fall 2024-2025

Continuous Assessment Test - I

Duration: 90 Mins

Program: B.Tech

Course Code: BCSE207L

Slot: G1**Course Name: Programming for Data Science**

Faculty: Dr. Sharmila Banu K (VL2024250101565) and Dr. Arunkumar T (VL2024250101558)

ANSWER ALL QUESTIONS

(5*10 = 50 Marks)

1a. Compare and contrast R and Python. Present a scenario to support your statements.

(5 Marks)

1b. Give a narrative of how Vectors can be used while handling Image or Audio or Tabular Data for respective Machine Learning Models (5 Marks)

2. The following table shows the features of 9 mobile phones introduced in the market about 6 months ago. A market research showed which of the mobile phones were liked by the public. Discuss the use of Vectors in designing a Logistic Regression for developing a model that can classify if a new phone will be liked by public or not.

(10 Marks)

Launch (within 6 months)	0	1	1	0	0	1	0	1	1
Weight (g)	151	180	160	205	162	182	138	185	170
Screen size (inches)	5.8	6.18	5.84	6.2	5.9	6.26	4.7	6.41	5.5
dual sim	1	1	0	0	0	1	0	1	0
Internal memory (≥ 64 GB, 4GB RAM)	1	1	1	1	1	1	1	1	1
NFC	0	1	1	0	1	0	1	1	1
Radio	1	0	0	1	1	1	0	0	0
Battery(mAh)	3060	3500	3060	5000	3000	4000	1960	3700	3260
Price (INR)	15k	32k	25k	18k	14k	12k	35k	42k	44k
Like (y)	1	0	1	0	1	1	0	1	0

3. Answer the following:

3a. Is it possible to create data of mixed types using Lists? (True/False) (2 Marks)

3b. Is it possible to load a set of images/Audio on to a list and display them?

Elaborate with example (5 Marks)

3c. Write the code to generate a sequence from 1 to 162 with an increment of 27.

Then repeat that sequence 5 times. Use system defined functions and highlight those (3 Marks)

4. The given dataset shows registration and CGPA details of 10 BCT Students.

(2 X 5 = 10 Marks)

Reg No	Credits Registered	Credits Earned	CGPA	Attendance	Bonus Marks Earned
22BCTXX01	83	83	8.45	86	0
22BCTXX02	76	69	9.41	95	1
22BCTXX03	83	83	9.01	60	1
22BCTXX04	83	40	3.38	64	0
22BCTXX05	83	83	8.11	75	2
22BCTXX06	85	85	8.77	77	1
22BCTXX07	77	73	7.36	78	1
22BCTXX08	79	76	7.96	65	1
22BCTXX09	80	80	9.09	50	0
22BCTXX10	83	83	7.55	90	2

Write separate functions for each of the following in R:

1. To find the highest CGPA
2. To find highest difference between credits registered and earned
3. To check if any CGPA is less than 4
4. To increase the Bonus marks by 1 if attendance > 80
5. To find the number of students whose CGPA > 7

5. A University is consulting with your Data Science Company for verifying the original writing of Graduate Applicants. They give the SoPs (Statement of Purpose) and GRE (Graduate Record Examination) English Essays of Candidates. Design a strategy to perform this task using the most suitable data structure in R and give recommendations. (10 Marks)

Hint: Similarity can be computed between two vectors

Note: Assume that ETS, the testing agency that conducts GRE agrees to share the Candidates' essay with the universities.
