

SVKM
Mithibai College (Arts, Sci & Comm)
Programme: B.Sc (Computer Science) - (CBCGS)
Year: III/Semester VI (Exam Year: 2023-2024)

Subject: DATA SCIENCE

Date: 03 Apr 2024

Max Marks: 75

Time: 10:30 am - 01:00 pm (02:30 Hrs.)

FINAL EXAMINATION(2023-2024)

Instructions:

1. This question paper contains 4 pages
2. Answer to each new question to be started on a fresh page.
3. Figure in right hand side indicates full marks
4. Assume suitable data if not given.
5. Use of Scientific calculator is allowed.

- | | |
|---|----|
| 1. (Attempt Any 3 Questions) | 15 |
| | |
| A. Discuss on Demographic data, Transactional data, Web behaviour data, Data from customer-created texts and spatial data | 5 |
| | |
| B. Demonstrate the need of IDE by considering following parameters: | 5 |
| 1) Features | |
| 2) definition | |
| | |
| C. Design KDD process for identifying knowledge from dataset. | 5 |
| | |
| D. According to you justify the importance of data management with respect to medical research(in case of COVID COVISHIELD development) | 5 |
| | |
| 2. (Attempt Any 3 Questions) | 15 |
| | |
| A. Develop python code to perform following | 5 |
| | |
| 1) Create three arrays with following specifications | |
| Marks = random numbers between 20 to 100 ; with 3 columns and 20 rows | |
| Sub=['eng','phy','chem'] | |
| Student names =[Manoj, Avishkar, Ashay, Raghav,] | |

- 2) Print complete information with the help of data frame
- 3) Compute $m1 = \text{eng} + \text{phy}$
- 4) Compute $m2 = \text{phy} + \text{chem}$
[assume suitable data if needed]
- B. Deepika is doing a research on cancer patients to predict whether person will have cancer in future or not. She wants to extract data from government hospital. How government hospital will show security concerns with respect to data? Explain in brief. 5
- C. Demonstrate working of numpy and pandas by giving comparative study 5
- D. Evaluate the following terms with respect to EDA 5
- Correlation Analysis
 - Handling Missing Data:
 - Descriptive Statistics
 - Identify Outliers:
 - Data Visualization:
3. (Attempt Any 3 Questions) 15
- A. Outline various functional and non-functional requirements of any research work associated with data science 5
- B. As a researcher propose the points why you want to use AWS datasets over other datasets? 5
- C. Use following points to prove various points associated with NoSQL Database: 5
- High scalability
 - High availability
 - Efficient big data management
 - Partition tolerance
 - Fast development
- D. Evaluate the role of RDF(resource description framework) by considering RDF triplets and example. 5
4. (Attempt Any 3 Questions) 15
- A. Compose application areas of logistic regression 5
- B. Justify role of lazy learners and Eager Learners 5

C. Consider the following data.

Id	Age	Competition	Type	profit
I1	old	Yes	s/w	Down
I2	Old	No	s/w	down
I3	Old	No	h/w	Down
I4	Mid	Yes	s/w	Down
I5	Mid	Yes	A/c	Down
I6	Mid	No	h/w	up
I7	Mid	No	s/w	Up
I8	New	Yes	A/c	Up
I9	New	No	h/w	Up
I10	new	no	s/w	Up

Compute information gain and Entropy and find root node

D. Consider following case study. There are 3 students A, B and C who appeared for mathematics exam. **Explain the cases given below and discuss and categorize them into overfitting, under fitting and best fit model by giving proper explanation or defining them.**

I) Student A, who was distracted in his own world, simply guessed the answers and got approximately 50% marks in the test.

II) Student B who memorized each and every question taught in the classroom was able to answer almost every question by memory and therefore obtained 98% marks in the class test.

III) For student C, she actually solved all the questions using the problem-solving approach she learned in the classroom and scored 92%

5. (Attempt Any 3 Questions)

A. Analyze the working of following python code and remove errors if any

```
arr = ['Animal', 'Thor', 'Agents', 'himalya']
ca = [word.capitalize for word in arr]
lo = [lower() for word in arr]
up = [word.upper() arr word ]
sw = [word.swapcase(arr) ]
ti = [word.title() for word in arr]
print("Original Array:", arr)
print("Capitalized:", ca)
print("Lowercased:", lo)
print("Uppercased:", up)
print("Swapcased:", sw)
print("Title-cased:", ti)
```

B. Examine the role of bar chart by writing python code to create name='a','b','c' and y=56,85,45.

5

C. Examine the role of data science process in research environment.

5

D. The values of x and their corresponding values of y are shown in the table below

5

x 0 1 2 3 4

y 2 3 5 4 6

Use linear regression to investigate the value of y when $x = 10$.