

No. of Pages: 1

Student Name: Kanis. H.

Enrolment No: 23

Mid-Term Examination – November 2022

Semester: Third, (Sept. 2022-Jan.2023)

Paper Name: Foundations of Data Science

Max. Marks: 30

Programme: B.Tech. (AI&DS, AI&ML, IOT)

Paper Code: AIDS 203/AIML 203/ IOT 203

Duration: 1.5 hrs.

Note:

- Question No. 1 is compulsory.
- Attempt any two questions from the remaining questions.
- Some questions have internal choice also.
- All questions carry equal marks.

Q. No.	Questions	Marks	BL	CO
		[2]	4	1
1(a)	Justify the statement "Information is a processed version of Data".	[2]	2	1
1(b)	Describe the significance of Python in the field of data science.	[2]	2	2
1(c)	Describe the difference between Bitwise left shift operator and Bitwise right shift operator with help of an example for a negative number.	[2]	2	2
1(d)	Compare and contrast the role of data scientist and data analyst.	[2]	2	2
1(e)	Explain Imputation. Demonstrate any two imputation techniques with help of examples.	[5]	4	1
2	Interpret the life cycle of data science with the help of a diagram and a case study highlighting the main procedural steps. Differentiate between data cleansing and data munging. OR a. Describe the significance of data understanding and domain knowledge with help of a case study.	[10] [5]	4 2	1
	b. Compare and contrast structured and unstructured data with real-world examples.	[5]	4	
3	Illustrate the process of identifying, handling, and replacing the missing value. State your assumptions and hypothesis for the same. OR a. Create a Python program for making a dynamic calculator.	[10] [5]	4 6	2
	b. What is a delimiter? Highlight the steps taken to import the data (CSV) from a given folder.	[5]	4	
4	Describe synthetic dataset and how is it different from natural datasets? Justify with an instance. How do you create a synthetic dataset in Python? Which packages and libraries are used for the same? OR Summarize the Data Cleaning Lifecycle with help of a real-life case study.	[10] [10]	3	

\*\*\*\*\*END\*\*\*\*\*

# END TERM EXAMINATION

THIRD SEMESTER [B.TECH] FEBRUARY 2023

Paper Code: AIDS/ AIML/ IOT203

Subject: Foundations of Data Science

Maximum Marks: 75

Time: 3 Hours

Note: Attempt five questions in all including Q.No.1 which is compulsory.  
Select one question from each unit.

Q1 Attempt all question

(3x5=15)

- a) What is data science? What is Seaborn and why is it a popular library for data visualization in Python?
- b) What is the purpose of clustering in data science and what are some common algorithms used for clustering?
- c) What is data exploration and why is it important in data science? What is the purpose of data normalization in data science?
- d) What is data cleaning and why is it necessary in the data science process? What is data visualization and what are its benefits in data science?
- e) What is data aggregation and how is it used in data science? What is data modeling and how is it different from data analysis?

## UNIT-I

- Q2 a) What makes Python a popular language for data science and machine learning? Describe. (5)
- b) How can you create a synthetic dataset in Microsoft Excel? (5)
- c) What are the most important traits of a successful data scientist? (5)
- Q3 a) What are some common techniques for processing unstructured data in data science? (5)
- b) What are the challenges in processing unstructured data in data science? (5)
- c) What are the primary responsibilities of a data scientist? What skills does a data scientist need? (5)

## UNIT-II

- Q4 a) How can you load and explore a dataset in Python using Pandas? (5)
- b) What are some common techniques for handling missing data in a dataset? (5)
- c) How do you handle exceptions in Python? (5)
- Q5 a) What is feature scaling and why is it important? How can we visualize the distribution of a numerical feature in a dataset in Python? (5)
- b) Can you explain the process of data analysis in Python using libraries such as pandas and numpy? (5)
- c) What are the different data types in Python and how do you declare a variable? (5)

P.T.O

AIDS/ AIML/ IOT-23  
P 1/2



[-2-]

### UNIT-III

- Q6 a) What are some popular libraries in Python for data science and machine learning and what are they used for? How can you install and manage packages in Python for data science? (7.5)
- b) How can you perform data cleaning and pre-processing in Python using Pandas and Numpy? (7.5)
- Q7 a) How can you perform statistical analysis in Python using Scipy and Statsmodels? (7.5)
- b) What are some common data visualization techniques in Python using Matplotlib and Seaborn? What are salient features of Matplotlib and Seaborn? (7.5)

### UNIT-IV

- Q8 a) What is machine learning and how can you perform supervised and unsupervised learning in Python using scikit-learn? (7.5)
- b) How can you evaluate the performance of a machine learning model in Python using metrics such as accuracy, precision, recall and F1-score? (7.5)
- Q9 a) What is trend analysis and how can it be performed in Python using Pandas and Matplotlib? (5)
- b) What is predictive mining and how can it be performed in Python using scikit-learn and other machine learning libraries? (5)
- c) What are the different types of recommendation algorithms used in recommender systems? (5)

\*\*\*\*\*

(Please Write Your Enrolment No. Immediately)

Student Name Konishk  
Enrolment No. 23

## Mid-Term Examination- JAN. 2023

Programme: B.Tech(AI&DS, AI&ML, IIOT)

Semester III

Paper Code: AIDS 207/AIML 207/IIOT 203

Subject: Foundation of Data Science

Time: 1.5 hours

Max. Marks: 30

Note:

- Q. No.1 is compulsory.
- Attempt any two questions from there maining questions.
- All Questions carry equal marks.
- Only a scientific calculator is allowed.

Question 1.		Marks	CO	BL
1(a)	Why do we need machine Learning in data science?	[2.5]	1	Understand
1(b)	Explain supervised and unsupervised machine Learning.	[2.5]	2	Understand
1(c)	What is the role of decision-making in data modeling?	[2.5]	2	Remember
1(d)	Write any four functions of Scipy Library.	[2.5]	3	Analyze

Question 2.		Marks	CO	BL
2(a)	What advantages does Numpy Array offer over nested Python List?	[5]	3	Understand
2(b)	Write a Python program for random number generation.	[5]	4	Remember

Question 3.		Marks	CO	BL
3(a)	Write code to design any recommender system.	[5]	4	Apply
3(b)	Discuss in detail Pandas in python with suitable examples.	[5]	3	Evaluate

Question 4.		Marks	CO	BL
4(a)	Explain the steps for the predictive model using python.	[5]	2	Understand Apply
4(b)	Describe plotting and visualization concepts in python.	[5]	3	Create



**Mid-Term Examination – November 2023**

Course: Foundation of Data Science

Paper Code: AIDS 203/AIML 203/ IOT 203

Programme: B.Tech. (AIDS, AIML, IOT)

Semester: Odd, Aug 2023-Jan 2024

Duration: 1.5 hrs.

Max. Marks: 30

**Note:**

- Question No. 1 is compulsory.
- Attempt any two questions from the remaining questions.
- All questions carry equal marks.

Q. No.	Question 1	Marks	CO
1(a)	Why is Python a popular programming language for data analysis and data science?	[2]	1
1(b)	Explain the difference between a data scientist and a data engineer?	[2]	1
1(c)	What is data science, and how does it relate to other disciplines like statistics and machine learning?	[2]	1
1(d)	Explain briefly the ethical considerations in data science, especially when handling sensitive data?.	[2]	1
1(e)	"How does data visualization enhance the initial exploration of a dataset, and can you provide a specific example illustrating its significance in uncovering insights from the data?"	[2]	2
<b>Question 2</b>			
2.	a. In your role as a data scientist, outline the specific steps you would take when confronted with highly unstructured data, lacking clear analysis guidelines. Describe the practical measures you'd apply to handle this situation effectively with the help of a case study.	[5]	1
	b. Investigate the fundamental knowledge domains related to data understanding and pre-processing within the realm of data analysis. Subsequently, present a comprehensive case study that demonstrates the substantial influence of employing data understanding and pre-processing techniques on the overall quality and results of a data analysis project.	[5]	2
<b>Question 3</b>			
3.	a. Describe the concept of synthetic dataset generation in the context of data science. Explain the reasons and circumstances where creating synthetic datasets becomes a necessity in data analysis.	[5]	2
	b. Analyze the importance of identifying and handling missing values in the field of data science. Discuss the strategies and methods for managing missing data effectively and their significance in ensuring data quality and meaningful analysis.	[5]	2
<b>Question 4</b>			
4.	a. Compare and contrast the efficiency of calculating the factorial of a number in Python using both recursive and iterative approaches.	[5]	3
	b. Analyze a real-world data science project in the finance sector. Provide a comprehensive overview of the data science problem addressed, the sources of data used, the data science methodologies and techniques applied, and the resulting impact on the business.	[5]	4

\*\*\*\*\*END\*\*\*\*\*



(Please write your Enrolment No Immediately)

Enrolment No. 62715611922

**CLASS TEST, DEC-2023**

**Programme : B.TECH. (AI&DS, AIML),**

**Semester 3<sup>RD</sup> SEM,**

**Paper Code: AIDS/AIML-203**

**Subject: Foundations of Data Science**

**Time: One Hour**

**Max. Marks : 30**

**Note: Q.No. 1 is Compulsory and any two more questions from remaining.**

Q. No1	a. Discuss various Functions used to create array in Numpy.	C03	2
	b. Discuss functions used to handle missing values in Pandas.	C03	2
	c. Differentiate Matplotlib and Seaborn in python.	C03	2
	d. What is clustering.	C04	2
	e. Discuss mathematical and scientific applications for Data Analysis.	C04	2
Q No 2.	a. How you can implement data processing in Pandas and Numpy.	C03	5
	b. Explain functions used in Matplotlib and Seaborn for visualization.	C03	5
Q. No 3.	a. Scipy has list of sub-packages , explain.	C03	5
	b. Difference between supervised and unsupervised learning.	C04	5
Q. No 4	a. What is predictive Mining? How it can be performed in python.	C04	5
	b. Explain Trend analysis, how it can performed in Python.	C04	5



# END TERM EXAMINATION

THIRD SEMESTER [B.TECH] JANUARY 2024

Paper Code: AIDS/AIML/IOT -203

Subject: Foundations of Data Science

Time: 3 Hours

Maximum Marks: 75

Note: Attempt five questions in all including Q. No.1 which is compulsory. Internal choice is indicated.

- Q1 Attempt all questions:-
- (a) Difference between data Analysis and analytics? (3)
  - (b) Explain supervised and unsupervised machine Learning? (3)
  - (c) What is data cleaning? Write the syntax to remove duplicate observations from a data frame in python? (3)
  - (d) Which are the best libraries for data visualization in python? Define IQR in a box plot. (3)
  - (e) How is np.mean() different from np.average() in NumPy? (3)
- Q2
- (i) Briefly explain the different steps of Data Science Process.
  - (ii) Why synthetic dataset generation needed? Give different commands in python for creating dataset.
  - (iii) Explain the process of data pre-processing. (15)
- Q3 Explain Jobroles of the following. (15)
- (i) Data Scientist.
  - (ii) Data Analyst
  - (iii) Machine learning expert.
  - (iv) Data engineer
  - (v) Data Architect
  - (vi) Data Administrator
  - (vii) Business Analyst.
  - (viii) Business Intelligence
- Q4
- (i) Explain in detail about data cleaning, integrating and transforming data in Data Science Process.
  - (ii) What do you mean by slicing operation in string of python? Write and example of slicing to fetch first name and last name from full name of a person and display it. (15)
- Q5 Write a python program to read a file. Illustrate the flow of the program. (15)
- Q6 Summarize the characteristics of NumPy, Pandas, SciPy and matplotlib libraries along with their usage in brief. Also, explain the typical methods to visualize data. (15)
- Q7 Write Python program that counts the words in its input and returns the most common ones. Also, Explain about Data frames. (15)
- Q8
- (i) Show the ways in which decision making and predictions are made in Data Science. Name some of the scientific methods used in data science.
  - (ii) Explain how a manufacturing industry using data science. (15)
- Q9 What is Recommender system? How you create a recommender system in python? Give and explain a real time example of recommender system. (15)

\*\*\*\*\*

AIDS/AIML/IOT-203  
P