P P SAVANI UNIVERSITY

Fifth Semester of B.Sc. IT Examination
DECEMBER 2021
SSIT3021 DATA SCIENCE

06.12.2021, Monday   Time: 9:00 a.m. To 11:30 p.m.   Maximum Marks: 60

Instructions:
1. The question paper comprises of two sections.
2. Section I and II must be attempted in separate answer sheets.
3. Make suitable assumptions and draw neat figures wherever required.
4. Use of scientific calculator is allowed.

SECTION – I

Q - 1	Answer the Following: (MCQ)	[05]
(i)	Which of the following is the most important language for Data Science?	

a) Java
b) Ruby
c) R
d) None of the mentioned

(ii) A different learning method does not include?
a) Memorization
b) Analogy
c) Deduction
d) Introduction

(iii) Which of the following is one of the key data science skills?
a) Statistics
b) Machine Learning
c) Data Visualization
d) All of the mentioned

(iv) Which of the following is tool for checking normality?
a). qqline()
b). qline()
c). anova()
d). lm()

(v) _Which of the following is used to find variance of all values?
a) var()
b) sd()
c) mean()
d) anova()

Q - 2 (a)	**Answer in Short.**	[05]
(i)	What do you mean by Median?	
(ii)	What is Data Science?	
(iii)	What is training data set?	
(iv)	What do you mean by Mode?	
(v)	What is the difference between primary data and secondary data?	
Q - 2 (b)	Describe types of data. Explain data types in brief.	[05]

OR

Q - 2 (b)	Distinguished between Structured and Unstructured Data.	[05]
Q - 3 (a)	Explain Naive Bayes classifier with an example.	[05]
Q - 3 (b)	Explain Linear regression with an appropriate example.	[05]

OR

Q - 3 (a)	Explain Support vector machine for linear classification.	[05]
Q - 3 (b)	What is Machine Learning? Explain each steps of Machine Learning.	[05]
Q - 4	Attempt any one	[05]
(i)	Explain Data Model? Explain various type of Data Model.	
(ii)	Enlist and explain each step of knowledge discovery from data.	

SECTION – II

Q - 1	Answer the Following: (MCQ/Short Question/Fill in the Blanks)	[05]
(i)	Which of the following thing can be data in Pandas?	

a) a python dict

b) an ndarray

c) a scalar value

d) all of the mentioned

(ii) Which of the following method make vector of repeated values?

a) rep()

b) data()

c) view()

d) read()

(iii) Which of the following finds the position of a quantile in a dataset?

a) quantile()

b) barplot()

c) barchart()

d} rep()

(iv) Data can be visualized using?

A. graphs

B. charts

C. maps

D. All of the above

(v) Which of the following input can be accepted by DataFrame?

a) Structured ndarray

b) Series

c) DataFrame

d) All of the mentioned

Q - 2 (a)	**Answer in Short.**	[05]
(i)	What do you mean by data encoding?	
(ii)	Differentiate Probability vs. Statistics	
(iii)	What is the difference between Data Scientist and Data Analytics?	
(iv)	Enlist two python libraries.	
(v)	What do you mean by visual encoding?	
Q - 2 (b)	Explain any one data analysis technique with an example.	[05]

OR

Q - 2 (b)	Explain application development methods used in data science.	[05]
Q - 3 (a)	Enlist and explain various data visualization techniques.	[05]
Q - 3 (b)	Explain data encoding techniques with an example.	[05]

OR

Q - 3 (a)	Explain retinal variables with suitable example.	[05]
Q - 3 (b)	Explain visual encoding technique.	[05]
Q - 4	Attempt any one	[05]
(i)	Explain any two applications of data science.	
(ii)	What do you mean by data visualization technology? Explain any one in detail.	

****

P P SAVANI UNIVERSITY

Fifth Semester of B.Sc. IT Examination
DECEMBER 2021
SSIT3021 DATA SCIENCE

06.12.2021, Monday   Time: 9:00 a.m. To 11:30 a.m.   Maximum Marks: 60

Instructions:
1. The question paper comprises of two sections.
2. Section I and II must be attempted in separate answer sheets.
3. Make suitable assumptions and draw neat figures wherever required.
4. Use of scientific calculator is allowed.

SECTION – I

Q - 1   **Answer the Following: (MCQ)**   [05]

(i)   Which of the following is performed by Data Scientist?
   a) Define the question
   b) Create reproducible code
   c) Challenge results
   d) All of the mentioned

(ii)   Machine learning is a subset of?
   a) Deep Learning
   b) Artificial Intelligence
   c) Data Learning
   d) None of the above

(iii)   What are the three types of machine learning?
   a) Supervised Learning
   b) Unsupervised Learning
   c) Reinforcement Learning
   d) All of the above

(iv)   Common classes of problems in machine learning is
   a) Clustering
   b) Classification
   c) Regression
   d) All of the above

(v)   What is Machine learning?
   a) The autonomous acquisition of knowledge through the use of computer programs
   b) The autonomous acquisition of knowledge through the use of manual programs
   c) The selective acquisition of knowledge through the use of computer programs
   d) The selective acquisition of knowledge through the use of manual programs

Q - 2 (a)   **Answer in Short.**   [05]
What are the components of Data Science?
What do you mean by Mean?
What is the difference between qualitative & quantitative Data?
What is testing data set?
What do you mean by Variance?

Q - 2 (b)   Explain Data Science process? Explain in Brief.   [05]

OR

Q - 2 (b)   Enlist and explain various data science toolkit.   [05]

Q - 3 (a)   Write a short note on SVM.   [05]

Q - 3 (b)   Explain Discrete probability distributions with their types.   [05]

Q - 3 (a)	Enlist and explain various machine learning algorithm.	[05]
Q - 3 (b)	Write a short note on KDD process.	[05]
Q - 4	Attempt any one	[05]
(i)	Enlist and explain multiple sources of data.	
(ii)	Explain data storage and its management methods.	

SECTION – II

Q - 1	Answer the Following: (MCQ/Short Question/Fill in the Blanks)	[05]
(i)	Which of the following is lattice command for producing a scatterplot?	
	a) plot()	
	b) lm()	
	c) xyplot()	
	d) anova()	
(ii)	Which of the following compute proportions from a contingency table?	
	a) par()	
	b) prop.table()	
	c) anova()	
	d) mosaicplot()	
(iii)	Which of the following is used to view dataset in a spreadsheet-type format?	
	a) Disp()	
	b) View()	
	c) Seq()	
	d) lm()	
(iv)	Which of the following groups find the correlation matrix?	
	a) factor.model	
	b) col.max(x)	
	c) stem	
	d) which.max(x)	
(v)	_____ function carries out a chi-square test.	
	a) chisq.test()	
	b) t.test()	
	c) prop.test()	
	d) fisher.test()	
Q - 2 (a)	**Answer in Short.**	[05]
(i)	What do you mean by data visualization technologies?	
(ii)	What is the difference between numpy and scipy?	
(iii)	What do you mean by Data Scientist?	
(iv)	What do you mean by Data Analyst?	
(v)	Differentiate keras and tensor flow.	
Q - 2 (b)	Explain any two data collection techniques.	[05]

OR

Q - 2 (b)	What are basic scales of measurement? Explain their types with example.	[05]
Q - 3 (a)	What is Data Visualization? Explain any one technique.	[05]
Q - 3 (b)	Write a short note on retinal variables.	[05]

OR

Q - 3 (a)	Enlist and explain any one data encoding techniques.	[05]
Q - 3 (b)	Define visual encoding. Explain it with suitable example.	[05]
Q - 4	Attempt any one	[05]
(i)	Enlist and explain various applications of data science.	
(ii)	Explain any two data visualization technologies.	

****