SRM VALLIAMMAI ENGINEERING COLLEGE

(An Autonomous Institution)

SRM Nagar, Kattankulathur – 603 203

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

QUESTION BANK

IV SEMESTER

1922403 – DATA SCIENCE CONCEPTS

Regulation – 2019

Academic Year 2021 – 2022 (EVEN SEMESTER)

Prepared by

Dr. S. JEYALAKSHMI, Assistant Professor (Sel. G) / AI & DS

SRM VALLIAMMAI ENGINEERING COLLEGE

(An Autonomous Institution)

SRM Nagar, Kattankulathur – 603 203.

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

QUESTION BANK

SUBJECT : 1922403 – DATA SCIENCE CONCEPTS

SEM / YEAR : IV Sem / II Year

UNIT I - INTRODUCTION			
Overview of Data Science: Challenges in Data Science, History of Data Science, Data Science Process, Discovery and Preparation, Model Planning and Building, Introduction to Python: Variables, Data types, Strings, Conditions and statements, Classes and objects - Type conversion, Functions and Packages			
PART – A			
1	What is Data Science?	BTL 2	Understanding
2	What are the job roles of Data Science?	BTL 4	Analyzing
3	What are the components of Data Science?	BTL 1	Remembering
4	What are the methods of collecting data?	BTL 1	Remembering
5	What is Data Science process? Explain.	BTL 2	Understanding
6	List out the areas in which Data Science can be applied.	BTL 2	Understanding
7	Who is a Data Scientist?	BTL 4	Analyzing
8	List the four main processes of Data Preparation	BTL 2	Understanding
9	Illustrate the use of Data Science with an example.	BTL 1	Remembering
10	Differentiate Data Mining and Data Science.	BTL 3	Applying
11	What is discovery in data Science?	BTL 6	Creating
12	Explain model planning and the tools available for model planning?	BTL 5	Evaluating
13	What are the sources of the data?	BTL 6	Applying
14	List out the phases of data science Life Cycle?	BTL 1	Remembering
15	List the standard data types in python.	BTL 3	Creating
16	List the various control flow structures	BTL 1	Remembering
17	List the types of Python operators.	BTL 1	Remembering
18	Define self parameter of Python	BTL 4	Analyzing
19	State the purpose of __init__.py file in Python.	BTL 3	Applying
20	Explain the arithmetic operators of Python	BTL 5	Evaluating

PART – B			
1	What do you understand by the term Data Science? Explain briefly the lifecycle of Data Science. (13)	BTL 5	Evaluating
2	List the Challenges of Data Science and Explain it in detail along with the solutions. (13)	BTL 1	Remembering
3	(i)Explain the importance of Data Science? (6) (ii) Describe the various applications of the data science? (7)	BTL 1	Remembering
4	Illustrate the various components of data science in detail? (13)	BTL 1	Remembering
5	Explain briefly about the data science process (13)	BTL 2	Understanding
6	Explain in detail the steps of Data Discovery and its benefits. (13)	BTL2	Understanding
7	Summarize the major tasks involved in data preparation. (13)	BTL 4	Analyzing
8	(i) Explain in detail the Top Tools for Data Scientists (6) (ii) Mention various Key features for Tools for Data Scientists (7)	BTL 2	Understanding
9	Illustrate the steps of data science model-building life cycle in detail. (13)	BTL 6	Creating
10	Write notes on following regarding Python variables (5+4+4) (i) Assigning multiple values (ii) Global variable (iii) output variable	BTL 3	Applying
11	Describe any 7 string methods of Python along with its syntax. (13)	BTL 1	Remembering
12	Explain the Logical, Identity, membership and bitwise operators of Python with syntax and example (13)	BTL 4	Analyzing
13	Summarize the conditional constructs of Python with syntax and example. Write notes on Python Functions (13)	BTL 4	Analyzing
14	Illustrate the Python Looping constructs in detail. (13)	BTL 2	Understanding
PART C			
1	State and explain the difference between data analytics and data science? (15)	BTL 5	Evaluating
2	How is Data Science different from traditional application programming?(15)	BTL 6	Creating
3	Discuss any five data science projects using API. (15)	BTL 6	Creating
4	Illustrate the step by step process of creating and executing Python package. (15)	BTL 5	Analyzing

UNIT II - DATA EXPLORATION AND PROCESSING			
Pandas - Data Structures, Series, DataFrame, NumPy - ndarray, SciPy - SciPy sub-packages, Data Structures, Matplotlib, Seaborn, Datashader. Data Processing: Processing CSV, JSON, XLS data, Data Wrandling, Data Aggregation			
PART – A			
Q.No	**Questions**	**BT Level**	**Competence**
1	List out the libraries in Python used for Data Analysis and Scientific Computations.	BTL1	Remembering
2	How will you create a series from a given list in Pandas?	BTL1	Remembering
3	What are the data structures in Pandas?	BTL1	Remembering

4	Differentiate Pandas series and Pandas Dataframe.	BTL3	Applying
5	What data structure is a Pandas dataframe?	BTL4	Analyzing
6	Are dataframes container for series?	BTL6	Creating
7	Define Pandas Panel	BTL3	Applying
8	What is NumPy Ndarray?	BTL1	Remembering
9	State the categories of array creation routines in NumPy.	BTL1	Remembering
10	Explain the purpose of zeros and ones functions in NumPy	BTL6	Creating
11	How to find the dimensions of the Ndarray?	BTL1	Remembering
12	What is SciPy?	BTL1	Remembering
13	In which language is SciPy written?	BTL3	Applying
14	Compare NumPy and SciPy.	BTL4	Analyzing
15	What is datashader?	BTL4	Analyzing
16	What are the types of output files in data processing?	BTL6	Creating
17	What is CSV file format?	BTL1	Remembering
18	Differentiate CSV and XLS format.	BTL2	Understanding
19	List some data wrangling tools.	BTL3	Applying
20	Define data aggregation.	BTL3	Applying
PART – B			
1	Summarize the different kinds of inputs accepted by dataFrame. (13)	BTL1	Remembering
2	Explain in detail the different ways to create Pandas dataframe. (13)	BTL1	Remembering
3	Illustrate all the basic operations that can be performed on Pandas series in detail. (13)	BTL2	Understanding
4	Describe in detail the Pandas series methods. (13)	BTL4	Analyzing
5	Write notes on (i) Methods for performing binary operations on Pandas (7) (ii)Compare Pandas and NumPy (6)	BTL4	Analyzing
6	Illustrate the various array creation routines for creating new array in NumPy along with its syntax. (13)	BTL5	Evaluating
7	Summarize the features of dtype object in NumPy Ndarray. (13)	BTL6	Creating
8	Explain the various sub packages in SciPy. (13)	BTL1	Remembering
9	Describe in detail the different kinds of plots with Matplotlib with example coding segment. (13)	BTL1	Remembering
10	(i) Compare Matplotlib and Seaborn (7) (ii) Benefits of data processing (6)	BTL1	Remembering
11	Explain with example (i) Parsing CSV files with Python's CSV library. (7)	BTL1	Remembering

	(ii) Parsing CSV files with the Pandas library (6)		
12	Write notes on the following while parsing xls file (i) Reading an excel file (5) (ii) Reading specific rows (4) (iii) Reading multiple excel files (4)	BTL5	Evaluating
13	Describe in detail the steps involved in data wrangling (13)	BTL2	Understanding
14	Explain the data aggregation and its importance in detail (13)	BTL4	Analyzing
PART C			
1	Describe in detail the techniques available in Pandas for Data manipulation with example. (15)	BTL6	Creating
2	Write a program using NumPy to (i) implement shaping and slicing in the array (8) (ii) find the maximum, minimum and sum of all elements in the array using NumPy Array Axis. (7)	BTL6	Creating
3	Discuss about the steps involved in processing Data and the methods of Data processing. (15)	BTL6	Creating
4	Summarize the operations of the following plotting functions with syntax (i) relplot (ii) lmplot (iii) catplot (5+5+5)	BTL6	Creating

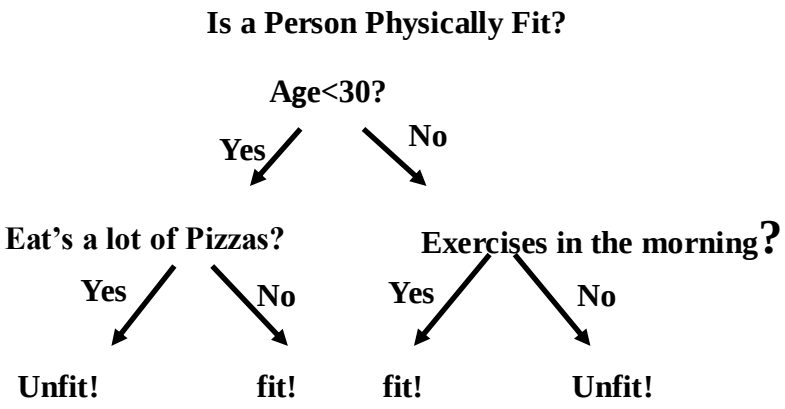
UNIT III – DATA VISUALIZATION

Context of data Visualization, Seven stages of Data Visualization, Objectives, Mapping, Chart Properties, Chart styling, Box plots, Heat Maps, Scatter Plots, Bubble Charts, 3D charts, Time Series, Geographical Data, Graph Data

PART – A

Q.No	Questions	BT Level	Competence
1	What is data visualization?	BTL 5	Evaluating
2	What are the types of data visualization?	BTL 5	Evaluating
3	State the pros and cons of data visualization.	BTL 1	Remembering
4	List some data visualization tools.	BTL 6	Creating
5	State the capabilities found in data visualization that meet the business needs.	BTL 4	Analyzing
6	List the Python libraries that are used for data visualization.	BTL 2	Understanding
7	Compare static and interactive visualization.	BTL 1	Remembering
8	Define time series visualization.	BTL 1	Remembering
9	Draw the diagram for interaction between the seven stages of data visualization.	BTL 6	Creating
10	What is meant by mapping?	BTL 2	Understanding
11	Explain the function to include annotations in chart.	BTL 2	Understanding
12	What are the different ways of drawing boxplot?	BTL 2	Understanding
13	Define heatmap.	BTL 1	Remembering

14	Illustrate scatterplot	BTL 4	Analyzing
15	State the methods used in heatmap and scatterplot.	BTL 3	Applying
16	Define bubble chart.	BTL 1	Remembering
17	What is the toolkit used for drawing 3D pot?	BTL 1	Remembering
18	Discuss time series and give example.	BTL 3	Applying
19	What is sparse matrix?	BTL 4	Analyzing
20	Which library is used in graph data?	BTL 3	Applying
	PART – B		
1	(i) Discuss the benefits of data visualization. (7) (ii) Explain in detail five important principles of data visualization (6)	BTL 1	Remembering
2	Explain in detail the seven stages of data visualization (13)	BTL 4	Analyzing
3	Summarize the interactions between the seven stages of data visualization with diagram (13)	BTL 1	Remembering
4	Write notes on (i) Chart properties (7) (ii) Chart styling (6)	BTL 1	Remembering
5	Illustrate data visualization in brief (13)	BTL 2	Understanding
6	Summarize the goals of data visualization (13)	BTL 5	Evaluating
7	Explain with example the box plot visualization (13)	BTL 6	Creating
8	Summarize the scatter plot method of data visualization with example code. (13)	BTL 2	Understanding
9	Explain the features of bubble chart in data visualization with example (13)	BTL 1	Remembering
10	Describe elaborately the 3D chart in data visualization. (13)	BTL 4	Analyzing
11	Illustrate time series chart with sample code segment. (13)	BTL 2	Understanding
12	Explain the geographical data chart with example. (13)	BTL 4	Analyzing
13	Explain sparse graph and the algorithms used to create sparse graph. (13)	BTL 3	Applying
14	Describe elaborately the heatmap with example code. (13)	BTL 3	Applying
	PART C		
1	Explain any three examples of data visualization. (15)	BTL 5	Evaluating
2	Summarize static visualization using any 2 real time example. (15)	BTL 6	Creating
3	Explain interactive visualization using any 2 real time example. (15)	BTL 5	Evaluating
4	Explain the graph data chart and the algorithms used in creation of sparse graph submodule. (15)	BTL 6	Creating

UNIT IV - MACHINE LEARNING			
Introduction to machine learning, Goals and applications of machine learning, aspects of developing a learning system, training data, concept representation, function approximation, Classification, Regression, Clustering, Decision trees, Recursive induction of decision trees			
PART – A			
Q.No	**Questions**	**BT Level**	**Competence**
1	Discuss Decision Tree algorithm	BTL 1	Remembering
2	What is pruning in a decision tree algorithm?	BTL 6	Creating
3	What is entropy in a decision tree algorithm?	BTL 3	Applying
4	Explain how Machine Learning is different from Deep Learning.	BTL 6	Creating
5	A common danger in machine learning is overfitting Justify	BTL 2	Understanding
6	What is Machine Learning?	BTL 1	Remembering
7	What is a model with respect to Machine Learning? Give example.	BTL 2	Understanding
8	How supervised models differ from unsupervised models?	BTL 4	Analyzing
9	Define reinforced learning.	BTL 3	Applying
10	Illustrate all possible decisions that can be made by the following decision tree. 	BTL 3	Applying
11	List the major categories of Machine Learning.	BTL 1	Remembering
12	Define simple linear Regression.	BTL 2	Understanding
13	What are the components of machine learning algorithm?	BTL 1	Remembering
14	Differentiate regression and Classification.	BTL 4	Analyzing
15	State the types of classification with example.	BTL 4	Analyzing
16	What is training data?	BTL 5	Evaluating
17	Discuss the types of regression.	BTL 5	Evaluating
18	List the types of supervised learning and unsupervised learning.	BTL 1	Remembering

19	What is the purpose of function approximation?	BTL 2	Understanding
20	What is a random forest? How does it work?	BTL 1	Remembering
	PART – B		
1	(i) Explain the differences between supervised and unsupervised learning.(6) (ii) How are Data Science and Machine Learning related to each other? (7)	BTL 1	Remembering
2	**Illustrate** decision trees with suitable examples. (13)	BTL 3	Applying
3	Discuss the various applications of machine learning. (13)	BTL 1	Remembering
4	Analyze how does machine learning works and the need for machine learning.	BTL 4	Analyzing
5	(ii) What are the primary trait of quality training data (7) (ii) Compare training data, test data and validation data (6)	BTL4	Analyzing
6	Discuss the types of clustering and the various clustering algorithms with example. (13)	BTL 2	Understanding
7	Discuss in detail the various Supervised Machine Learning techniques. (13)	BTL 3	Applying
8	Describe briefly the strengths and weaknesses of decision tree approach. (13)	BTL 1	Remembering
9	Illustrate the seven major steps involved in the life cycle of machine learning.	BTL 2	Understanding
10	Discuss random forest with suitable algorithms and examples? (13)	BTL 1	Remembering
11	Explain the three main types of regression with example. (13)	BTL 5	Evaluating
12	Write routine for logistic regression and explain with necessary data and charts. (13)	BTL 6	Creating
13	Differentiate classification model and regression model of machine learning with suitable examples. (13)	BTL 4	Analyzing
14	Discuss the applications of clustering. (13)	BTL 2	Understanding
	PART C		
1	**Construct** a decision tree for the following data: Explain various path in the tree that leads to various decisions. (15)	BTL 5	Evaluating

Outlook	**Temp**	**Humidity**	**Windy**	**Play Golf**
Rainy	Hot	High	False	No
Rainy	Hot	High	True	No
Over roast	Hot	High	False	Yes
Sunny	Mild	High	False	Yes
Sunny	Cool	Normal	False	Yes
Sunny	Cool	Normal	True	No
Over roast	Cool	Normal	True	Yes
Rainy	Mild	High	False	No
Rainy	Cool	Normal	False	Yes
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	True	Yes
Overroast	Mild	High	True	Yes
Overroast	Hot	Normal	False	Yes
Sunny	Mild	High	True	No

2	Evaluate random trees and explain random forest. (15)	BTL 6	Creating
3	Construct a decision tree for sample data of your own and evaluate various decision that can be arrived based on the decision tree. (15)	BTL 6	Creating
4	Summarize the types of classification, the algorithms used for each classification and explain with example. (15)	BTL 5	Evaluating

UNIT V - DEEP LEARNING

Introduction to Deep Learning, Supervised Learning With Neural Networks, Neural Network Basis: Binary Classification, Logistic Regression, Derivatives, Computation graph, Broadcasting in python, Neural Style transfer, text generation and image generation

PART – A

Q.No	Questions	BT Level	Competence
1	What is logistic regression in Data Science?	BTL1	Remembering
2	What is Deep Learning?	BTL1	Remembering
3	Why is it called deep learning?	BTL1	Remembering
4	What is supervised learning in Neural network?	BTL1	Remembering
5	What are three main categories of Neural network?	BTL1	Remembering
6	What is the difference between CNN and RNN?	BTL1	Remembering
7	What are the basic elements of perceptron?	BTL2	Understanding
8	Define back propagation.	BTL2	Understanding
9	State the steps involved in the process of CNN.	BTL2	Understanding
10	State the steps involved in the process of RNN.	BTL2	Understanding
11	What is firing of a neuron?	BTL3	Applying
12	What is forward propagation?	BTL3	Applying
13	State the important features of Adaptive Linear Neuron.	BTL3	Applying
14	Analyze the important features of Multiple adaptive linear neurons.	BTL4	Analyzing
15	Define derivative in deep learning.	BTL4	Analyzing
16	State the reason for using derivatives of a activation functions in a neural network.	BTL4	Analyzing
17	What is a computational graph in deep learning?	BTL5	Evaluating
18	Explain the broadcasting feature in NumPy.	BTL5	Evaluating
19	Where is neural style transfer used?	BTL6	Creating
20	Define text generation.	BTL6	Creating

PART-B

1	(i) Differentiate Deep learning and Machine learning. (7) (ii) Explain the applications of Deep learning. (6)	BTL1	Remembering
2	Explain the various classifications of Deep learning algorithms in detail. (13)	BTL1	Remembering
3	(i) Illustrate with neat diagram the model of Artificial Neural Network. (7) (ii) How is the structure of neural network determined? (6)	BTL1	Remembering
4	(i) How do deep learning algorithms learn? (4) (ii) Explain back propagation neural network with its architecture diagram. (9)	BTL1	Remembering
5	Explain in detail the adaptive linear neuron architecture and training algorithm. (13)	BTL2	Understanding
6	Explain briefly how the logistic regression links with neural network. (13)	BTL2	Understanding
7	Write notes on Computational graphs and back propagation (13)	BTL2	Understanding
8	Describe elaborately computational graphs with example (13)	BTL3	Applying
9	Write notes on (i) Rules of broadcasting (7) (ii) Limitations of broadcasting (6)	BTL3	Applying
10	Illustrate the features of Broadcasting in NumPy (13)	BTL4	Analyzing
11	Discuss Neural style transfer and its implementation. (13)	BTL4	Analyzing
12	Explain text generation and image generation in detail (13)	BTL4	Analyzing
13	Criticize logistic regression in detail. (13)	BTL5	Evaluating
14	Elaborate a brief note on supervised learning with neural networks. (13)	BTL6	Creating
PART C			
1	Illustrate binary classification using CNN model in detail. (15)	BTL5	Evaluating
2	Explain the types of activation function and describe under which condition those activation functions may be used? (15)	BTL6	Creating
3	Describe in detail the steps involved in perceptron training algorithm with architecture diagram for (i) Single output unit (7) (ii) Multiple output unit (8)	BTL6	Creating
4	Illustrate the components of multiple adaptive linear neuron architecture and the training algorithm. (15)	BTL5	Evaluating