U.S.N | | | | | | | | | | |

P.E.S. College of Engineering, Mandya - 571 401

(An Autonomous Institution affiliated to VTU, Belagavi)

Seventh Semester, B.E. - Information Science and Engineering

Semester End Examination; February - 2022

Data Science

Time: 3 hrs *Max. Marks: 100*

Course Outcomes

The Students will be able to:

CO1: Explain Data Science process and Statistical Inference.

CO2: Apply basic tools (plots, graphs, summary statistics) to carry out EDA and identify basic Machine Learning algorithms to use in applications.

CO3: Use APIs and other tools to scrap the Web and identify basic Feature Generation and Feature Selection algorithms to use in applications.

CO4: Build own recommendation system.

CO5: Create effective visualization of a given data (to communicate or persuade ethically).

***Note**: I) **PART - A** is compulsory. **Two** marks for each question.*

*II) **PART - B**: Answer any **Two** sub questions (from a, b, c) for Maximum of **18 marks** from each unit.*

Q. No.		Questions	Marks	BLs	COs	POs
		I : PART - A	**10**			
I	a.	Represent Data Science using Drew Conway's Venn diagram.	2	L1	CO1	PO1
	b.	Why and when do we perform linear regression?	2	L1	CO2	PO1
	c.	What are wrappers? List the two aspects we need to consider while using wrapper for feature selection.	2	L1	CO3	PO1
	d.	What is curse of dimensionality? Which are the measures to solve it?	2	L1	CO4	PO1
	e.	List any four examples of inspiring (industry) projects for data visualization.	2	L2	CO5	PO1
		II : PART - B	**90**			
		UNIT - I	**18**			
1	a.	Discuss the reasons for growing popularity of data science in recent times. Also, explain the concept of datafication.	9	L2	CO1	PO1
	b.	Explain the following: i) Statistical Inference ii) Populations and Samples	9	L2	CO1	PO1
	c.	Explain the following assumptions with respect to big data domain: i) Incomplete assumptions ii) Other bad or wrong assumptions	9	L2	CO1	PO1
		UNIT - II	**18**			
2	a.	With a neat diagram, illustrate the process of data science and involvement of data scientist in the process.	9	L2	CO2	PO1

Contd… 2

	b. Explain prediction using k-NN algorithm. Also, discuss various similarity measures used in k-NN.	9	L2	CO2	PO1
	c. Explain simple linear regression equation and multiple linear regression equation. Also, discuss the evaluation metrics. For the following dataset, consider $\beta_0 = 0.4$ and $\beta_1 = 0.8$. Find the predicted y for each of the instance.	9	L2	CO2	PO1

1	x	y
2	1	1
3	2	3
4	4	3
5	3	2
6	5	5

UNIT - III **18**

3 a.	Explain Navie Bayes algorithm and also explain the purpose of Laplace smoothing.	9	L3	CO3	PO2
	b. Explain the different types of feature selection methods (Stepwise regression) and selection criterions.	9	L2	CO3	PO1
	c. Illustrate the importance of entropy in constructing a decision tree. Also, describe the process of constructing decision tree.	9	L2	CO3	PO1

UNIT - IV **18**

4 a.	Explain, why k-NN algorithm is not suitable for recommendation engines?	9	L2	CO4	PO1
	b. Write a note on Singular Value Decomposition.	9	L2	CO4	PO1
	c. Discuss the types of social networks and the centrality measures used in social networks.	9	L2	CO4	PO1

UNIT - V **18**

5 a.	List out and explain any four benefits of data visualization.	9	L2	CO5	PO1
	b. With a suitable example, discuss the detection of suspicious activity using data visualization.	9	L1	CO5	PO1
	c. Explain the following with respect to data science: i) Privacy ii) Security iii) Ethics	9	L2	CO5	PO1

* * * *