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 / March - 2023
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).

<u>*Note:*</u> *I) PART - A is compulsory. **Two** marks for each question.*
  *II) PART - B: Answer any <u>**Two**</u> 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**			
1 a.	What is datafication? Give an example.	2	L1	CO1	PO1
b.	What are the three basic classes of algorithm a data scientist should be aware of?	2	L1	CO2	PO1
c.	What is feature selection?	2	L1	CO3	PO1
d.	Write any two essential characteristics of a social network.	2	L1	CO4	PO1
e.	List any four examples of inspiring projects for data visualization.	2	L1	CO5	PO1
	II : PART - B	**90**			
	UNIT - I	**18**			
2 a.	Describe data science and the skill sets needed for a data scientist.	9	L2	CO1	PO1
b.	Is the role of data scientist in academia and industry same? Justify.	9	L2	CO1	PO1
c.	What is a model? Explain how a model is built and fit, also what is the meaning of over fitting of a model?	9	L2	CO1	PO1
	UNIT - II	**18**			
3 a.	With a neat diagram, discuss the data science process and the role of data scientist in this process.	9	L2	CO2	PO1
b.	How can the linear regression be extended beyond least square method? Discuss the same in detail.	9	L2	CO2	PO1
c.	With a suitable scenario, discuss k-nearest neighbour for classification task.	9	L2	CO2	PO1

	UNIT - III	**18**			
4 a.	Why linear regression and k-Nearest Neighbor are poor choice for spam filtering problem? Discuss the same.	9	L2	CO3	PO1
b.	What are the three categories of feature selection methods? Discuss the same.	9	L2	CO3	PO1
c.	Discuss the working of Decision Tree algorithm.	9	L2	CO3	PO1
	UNIT - IV	**18**			
5 a.	Discuss the problems associated with Nearest Neighbor's algorithm for building recommendation engines.	9	L2	CO4	PO1
b.	Compare and contrast SVD and PCA.	9	L3	CO4	PO1
c.	Explain any three techniques used for discovering communities directly in a social network.	9	L2	CO4	PO1
	UNIT - V	**18**			
6 a.	List and explicate the benefits of data visualization.	9	L2	CO5	PO1
b.	List and discuss the tools used for data visualization.	9	L2	CO5	PO1
c.	Explain the following with respect to data science: i) Privacy ii) Security iii) Ethics	9	L2	CO5	PO1

∗ ∗ ∗