Time: 3 Hours   **Marks: 80**

Note: 1. Question 1 is compulsory

2. Answer any three out of remaining questions.

3. Assume suitable data wherever required and justify the same.

Q1	a)	Explain the Importance of Ethics and related issue in Data Science.	[5]
	b)	What are the differences between supervised learning and unsupervised learning, elaborate with an example.	[5]
	c)	What is the background of the formation of data journalism?	[5]
	d)	How bootstrap method can be used for evaluating the accuracy of a classifier.	[5]

Q2   a) Use PCA to transform 2D data space to 1D data space for the given matrix A.   [10]

$$A = \begin{matrix} 0 & 1 \\ -2 & -3 \end{matrix}$$

b) Explain Maximum Likelihood estimation using expectation–maximization (EM).   [10]

Q3   a) Discuss how the ROC curve can be used to determine an appropriate threshold value for a classifier.   [10]

b) What is Data Science? How is it different from Big Data and Data Analytics?   [10]

Q4   a) What type of problem were you looking to solve with text mining? How did you know how to text mine? What could be the challenges when text mining?   [10]

b) How Sentiment Analysis can be used for Social Media Monitoring? Elucidate with suitable case.   [10]

Q5   a) Draw and describe the information visualization process.   [10]

b) List and explain basic relational operators used in pig with example.   [10]

Q6   a) What infrastructure is most appropriate for Hadoop? Draw and describe Hadoop Ecosystem Architecture.   [10]

b) What is the big deal with Big Data? If your company is just starting to consider using Big Data in you marketing research, what would be most useful to include? Explain in Data Science perspective.   [10]

F691584A4BE2D03B98D51BFBE1AA595E