

## Sub: Data Science

### Question Bank on Unit 3,4 ,5 and 6

Sr. No	Question	Marks
<b>Unit 3 Data Analytics Life Cycle</b>		
1	Explain Data Analytics life cycle with the help of diagram	10
2	List different phases in data analytics life cycle and explain Model Building phase in detail.	8
3	What are different phases in data analytics life cycle? Explain Operationalize phase in detail	10
4	Explain Model building phase with its challenges	8
5	List and explain the steps in data preparation phase of data analytics life cycle.	8
6	Write short note on the following: i) ETL ii) Common tools for the model building. iii) Model selection for data analytics	9
7	Explain model planning in detail	6
8	Explain operationalize phase in details	6
9	Explain discovery phase in detail	6
10	Write a note on Communication results	6
<b>Unit 4 Predictive Data Analytics with Python</b>		
1	Explain association rules with example.	4
2	Explain Python Libraries for Data Processing, Modeling and Data Visualization.	10
3	Explain predictive, Descriptive, and Prescriptive data analysis. And also mention their difference.	6
4	What are the types of analytics in big data? Explain in brief.	9
5		9

	Calculate the support and confidence value for all the possible item sets.													
	<table><tr><th>Transaction ID</th><th>Items bought</th></tr><tr><td>1</td><td>Onion, Potato, Cold drink</td></tr><tr><td>2</td><td>Onion, Burger, Cold drink</td></tr><tr><td>3</td><td>Eggs, Onion, Cold drink</td></tr><tr><td>4</td><td>Potato, Milk, Eggs.</td></tr><tr><td>5</td><td>Potato, Burger, cold drink, Milk eggs.</td></tr></table>	Transaction ID	Items bought	1	Onion, Potato, Cold drink	2	Onion, Burger, Cold drink	3	Eggs, Onion, Cold drink	4	Potato, Milk, Eggs.	5	Potato, Burger, cold drink, Milk eggs.	
Transaction ID	Items bought													
1	Onion, Potato, Cold drink													
2	Onion, Burger, Cold drink													
3	Eggs, Onion, Cold drink													
4	Potato, Milk, Eggs.													
5	Potato, Burger, cold drink, Milk eggs.													
6	Explain the use of logistic function in logistic regression in detail	9												
7	Write short note on the following: i) Removing duplicates from data set. ii) Handling missing data iii) Data transformation	9												
8	Explain why decision tree are used. Draw a sample decision tree and explain its parts	9												
9	How Apriori Algorithm works, explain with suitable example?	9												
10	Explain Naïve Bayes' classifier and it applications.	9												
11	What is data preprocessing? Explain in details about handling missing data and transformation of data.													
12	What is scikit-learn library? Explain its usage	6												
13	Explain FP growth algorithm with example	6												
14	Explain linear regression with example	6												
Unit 5 Data Analytics and Model Evaluation														
1	What is clustering? With suitable example explain the steps involved in k-means algorithm.	7												
2	Discuss Holdout method and Random Sampling methods	6												
3	Wirte short note on i) Confusion matrix ii) AVC- ROC curve	8												
4	What do you mean by text analysis? Why text analysis need to be done? Explain the following text analysis steps with suitable examples i) Part of speech (POS) tagging ii) Lemmatization iii) Stemming	11												
5	Write short note on i) Time series Analysis ii) TF- IDF	6												
6	Suppose that the given data the taste is to cluster points (With (x.y) representing location) into three cluster, where the points are. A1(2,10), A2(2,5), A3(8,4),	9												

	B1 (5,8) B2(7,5) B3(6,4), C1(1,2), C2(4,9) The distance function is Euclidean distance suppose initially we assign A1, B1 and C1 as the center of each cluster, respectively. use the kmeans algorithm to show only the three cluster centers after the first round of execution with steps.														
7	Explain the following text analysis steps with suitable example. i) Part of speech (POS) tagging ii) Lemmatization iii) Stemming	8													
8	a) Given the confusion matrix, calculate accuracy, precision, Recall, Error rate with description on heart attack risk. <table><tr><td colspan="2" rowspan="2"></td><th colspan="2">Predicted classes</th></tr><tr><th>Heart-Attack Risk-yes</th><th>Heart Attack Risk-No</th></tr><tr><th rowspan="2">Actual Classes</th><th>Heart Attack Risk-yes</th><td>80</td><td>220</td></tr><tr><th>Heart Attack Risk-No</th><td>150</td><td>9,500</td></tr></table>			Predicted classes		Heart-Attack Risk-yes	Heart Attack Risk-No	Actual Classes	Heart Attack Risk-yes	80	220	Heart Attack Risk-No	150	9,500	8
				Predicted classes											
		Heart-Attack Risk-yes	Heart Attack Risk-No												
Actual Classes	Heart Attack Risk-yes	80	220												
	Heart Attack Risk-No	150	9,500												
9	Explain the TF/IDF (term frequency-inverse document frequency) terms in text analysis with suitable example	9													
10	Write a note on Parameter Tuning and Optimization	4													
11	Define following terms with respect to confusion matrix : i) Accuracy ii) Precision iii) Recall iv) AUC-ROC	8													
12	Explain k-fold Cross Validation & Random Subsampling	9													
13	With suitable example ,explain the steps involved in k-means algorithm	9													
14	What is text processing? Explain TF-IDF with example	8													
15	Write a note on social network analysis	6													
Unit 6 Data Visualization and Hadoop															
1.	With a suitable example explain Histogram and explain its usages	6													
2.	Explain in detail the Hadoop Ecosystem with suitable diagram.	6													
3.	Describe the Data visualization tool “Tableau”. Explain its applications in brief	6													
4.	What are the major challenges in visualizing the big data and how to over	6													

	come these challenges.	
<b>5.</b>	List and explain 4 different tools for data visualization.	<b>6</b>
<b>6.</b>	Write a note on: i) Pie charts ii) Bar graphs iii) Scatterplots	<b>9</b>
<b>7.</b>	Explain different Customization of plots and Graphical representations used in Data Science.	<b>6</b>
<b>8.</b>	Explain Data Visualization with Tableau	<b>4</b>
<b>9.</b>	Explain different data visualization tools.	<b>6</b>
<b>10.</b>	Explain any 4 Types of data visualization with example.	<b>8</b>
<b>11.</b>	Explain Role Job tracker and Task Tracker in Hadoop Architecture.	<b>5</b>
<b>12.</b>	Write 5 Hadoop Shell commands.	<b>5</b>
<b>13.</b>	Explain Map Reduce with proper diagram for word count example.	<b>8</b>
<b>14.</b>	Draw and explain Architecture of HIVE	<b>6</b>
<b>15.</b>	Explain architecture of Apache-Pig.	<b>6</b>
<b>16.</b>	What is Data Visualization? Explain different techniques of data visualization in detail.	<b>6</b>
<b>17.</b>	b) With proper python code, draw and explain. i) Line plot ii) Scatter plot iii) Box plot iv) Histogram	<b>6</b>
<b>18.</b>	Explain analytical techniques used in big data visualization.	<b>6</b>

<b>19.</b>	Describe the workflow between Hive and Hadoop with suitable diagram	<b>6</b>
<b>20.</b>	What is data visualization? What are the different methods of data visualization explain in detail.	<b>6</b>
<b>21.</b>	What are the different types of data visualization?	<b>6</b>
<b>22.</b>	Write a note on Apache Pig and Apache Hive	<b>6</b>