



Mahavir Education Trust's  
**SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE**  
Chembur, Mumbai - 400 088  
**UG Program in Information Technology**

Sample paper for DA ME 2020-21

Q1.	Choose the correct option for following questions. All the Questions are compulsory and carry equal marks
1.	Structured data resides in a _____ field within a record.
Option A:	Fixed
Option B:	Varying
Option C:	Changing
Option D:	Image
2.	Which of the following is not a characteristic of big data?
Option A:	View
Option B:	Volume
Option C:	Variety
Option D:	Velocity
3.	Data in Data Science is _____
Option A:	Structured and Unstructured Data
Option B:	Structured and Unstructured Data
Option C:	Unstructured Data only
Option D:	Time Series Data only
4.	Business in Data Science is _____
Option A:	Analytics-Owned
Option B:	IT-Owned
Option C:	Methodical
Option D:	Performance Management
5.	Tools of Data Science are _____
Option A:	Cloud Centric
Option B:	Relational Database Centric
Option C:	Update Centric
Option D:	Backup Specific
6.	Which of the following is a Data Scientist analysis task?
Option A:	Prescriptive
Option B:	Retrospective
Option C:	Descriptive
Option D:	Enumerative
7.	Which process is not followed by a Data Scientist?
Option A:	Static
Option B:	Exploratory



Mahavir Education Trust's  
**SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE**  
 Chembur, Mumbai - 400 088  
**UG Program in Information Technology**

Option C:	Experimentation
Option D:	Visual
8.	Which of the following is Data Scientists' Area of focus?
Option A:	Correlation
Option B:	Reports
Option C:	KPIs
Option D:	Trends
9.	How could the matrix be constructed by using the following R code? <code>m &lt;- matrix(1:6, nrow = 2, ncol = 3)</code>
Option A:	row-wise
Option B:	column-wise
Option C:	any manner
Option D:	data insufficient
10.	The datanode and namenode are respectively
Option A:	Master and worker nodes
Option B:	worker and Master nodes
Option C:	co-locate the data with the computing nodes.
Option D:	all masters
11.	The expected value or _____ of a random variable is the center of its distribution.
Option A:	Median
Option B:	Mean
Option C:	Mode
Option D:	Bayesian Inference
12.	Which function is used to create the vector with more than one element?
Option A:	<code>plot()</code>
Option B:	<code>c()</code>
Option C:	<code>par()</code>
Option D:	<code>library()</code>
13.	Which function takes a <code>dim</code> attribute which creates the required number of dimensions?
Option A:	Vector
Option B:	Lists
Option C:	Array
Option D:	Matrix



Mahavir Education Trust's  
**SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE**  
Chembur, Mumbai - 400 088  
**UG Program in Information Technology**

14.	SVM find a _____ for separation of the data using essential training tuples called support vectors
Option A:	hyperplane
Option B:	line
Option C:	mean
Option D:	distribution
15.	Model-based clustering include the _____ algorithm
Option A:	B tree
Option B:	Expectation-Maximization
Option C:	Decision tree
Option D:	Lazy learner
16.	Which of the following is not a text analysis steps,
Option A:	tokenization
Option B:	stemming
Option C:	Run-length encoding
Option D:	lemmatization
17.	Document clustering is _____ approach
Option A:	Supervised
Option B:	Unsupervised
Option C:	Randomized
Option D:	Semi-supervised
18.	Which of the following is Lazy Learner
Option A:	K Nearest Neighbor
Option B:	Decision Tree Classifier
Option C:	Random Forest Classifier
Option D:	Adaboost Classifier
19.	Which of the following is Eager Learner
Option A:	B tree
Option B:	Expectation-Maximization
Option C:	Decision tree
Option D:	Lazy learner
20.	Data Frames are created using the?
Option A:	frame() function
Option B:	data.frame() function
Option C:	data() function
Option D:	frame.data() function



Mahavir Education Trust's  
**SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE**  
Chembur, Mumbai - 400 088  
**UG Program in Information Technology**

<b>Q2</b>	<b>Solve any Two Questions out of Three</b>	<b>10 marks each</b>
A	Discuss any Business problems and data science solutions towards it	
B	Compare model based clustering algorithms	
C	Explain k-nearest neighbor algorithm	

<b>Q3.</b>	<b>Solve any Four out of Six</b>	<b>5 marks each</b>
A	List and explain functions for EDA in R	
B	Compare statistical methods for evaluation	
C	Explain data visualisation types	
D	Illustrate interactive dashboards with examples	
E	Explain text analysis steps,	
F	Explain Data Science Process	