

Roll Number: _____ IIPS DAVV M.Tech. 5th Sem (Data Science)-Test-1

Q.1: Discuss, attribute by attribute, problems that may have found into following database:

CustID	Zip	Gender	Income	Marital Status	Transaction Amount
1001	10048	M	75000	M	5000
1002	J2S7K7	F	-40000	W	4000
1003	90210		10000000	S	7000
1004	6269	M	50000	S	1000
1005	55101	F	99999	D	3000

Q.2: Apply data wrangling process in following database:

Name	Phone	Birth date	State
Jadish, Singh	445-881-4478	August 12, 1979	Maharashtra
Jain Tilak	+91-189-456-4513	11/12/1975	TN
Gupta, Binod	(731)546-8165	June 15, 82	Karnataka
Arun Fakira	5493156648	2-6-1985	Assam
Jackie Agrawal	256-4896	January 30	Alabama

Q.3: Consider following database, apply any 3 methods for finding out missing values:

ID	A1	A2	A3	A4	- With user defined constants - With Mean or Mode - With Random draws from the distribution of the variables
1	14.0	350	165	Ujjain	
2	31.9		71	Etanagar	
3	17.0	302	140	Ujjain	
4	15.0	400	150		
5	37.7	89	62	Jaipur	

1.

Q.4: Classify bank loan applications by assigning applications to one of the 3 risk class:
(Use Entropy and Information Gain approach)

Home?	Married?	Gender?	Employed?	Credit rating?	Risk class?
Y	Y	M	Y	A	B
N	N	F	Y	A	A
Y	Y	F	Y	B	C
Y	N	M	N	B	B
N	Y	F	Y	B	C
N	N	F	Y	B	A
N	N	M	N	B	B
Y	N	F	Y	A	A
N	y	F	Y	A	C
y	y	F	Y	A	C

Q.5: Compare the attributes for classification of bank loan applications by
Assigning applications to one of the 3 risk class: (Use Gini Index and Information gain approach)

Roll Number : _____ IIPS DAVV M.Tech. 5th Sem (Data Science)-Test-1

Q.1: Discuss, attribute by attribute, problems that may have found into following database:

CustID	Zip	Gender	Income	Marital Status	Transaction Amount
1001	10048	M	75000	M	5000
1002	J2S7K7	F	-40000	W	4000
1003	90210		10000000	S	7000
1004	6269	M	50000	S	1000
1005	55101	F	99999	D	3000

Q.2 : Apply data wrangling process in following database :

Name	Phone	Birth date	State
Jagdish, Singh	445-881-4478	August 12, 1979	Maharastra
Jain Tilak	+91-189-456-4513	11/12/1975	TN
Gupta, Binod	(731)546-8165	June 15, 82	Karnatak
Arun Fakira	5493156648	2-6-1985	Asam
Jackie Agrawal	256-4896	January 30	Alabama

Q.3: Consider following database, apply any 3 methods for finding out missing values:

ID	A1	A2	A3	A4	- with user defined constants - With Mean or Mode - With Random draws from the distribution of the variables
1	14.0	350	165	Ujjain	
2	31.9		71	Etanagar	
3	17.0	302	140	Ujjain	
4	15.0	400	150		
5	37.7	89	62	Jaipur	

Q.4: Classify bank loan applications by assigning applications to one of the 3 risk class:
(Use Entropy and Information Gain approach)

Home?	Married?	Gender?	Employed?	Credit rating?	Risk class?
Y	Y	M	Y	A	B
N	N	F	Y	A	A
Y	Y	F	Y	B	C
Y	N	M	N	B	B
N	Y	F	Y	B	C
N	N	F	Y	B	A
N	N	M	N	B	B
Y	N	F	Y	A	A
N	y	F	Y	A	C
y	y	F	Y	A	C

Q.5 : Compare the attributes for classification of bank loan applications by
assigning applications to one of the 3 risk class: (Use Gini Index and Information gain approach)