1

Perform data preprocessing on following dataset.

-Analyses on the dataset
-preprocess on null dataset
-what is dirty data set
-how to handle continues values
-differentiate between continues and discrete/categorical values.

Owns Home	Married	Gender	Employed	Creadit Rating	Risk Class
Yes	Yes	Male	Yes		B
No	No	Female	Yes		A
Yes	Yes	Female	Yes	B	C
Yes		Male	No	B	B
No	Yes	Female	Yes	B	C
No	No	Female	Yes	B	A
No	No	Male	No	B	B
Yes		Female	Yes	A	A
No	Yes	Female	Yes	A	C
Yes	Yes	Female	Yes	A	C

2

Perform data preprocessing on following dataset.

-Analyses on the dataset
-preprocess on null dataset
-what is dirty data set
-how to handle continues values
-differentiate between continues and discrete/categorical values.

Sr.No	Debt	Collateral	Income	Credit-History	Risk
1	high	none	0-15k	bad	high-risk
2	high	none	15-35k	unknown	high-risk
3	low	none	15-35k	unknown	Moderate risk
4	low	none	0-15k		high-risk
5	low	none	over 35k		low risk
6	low	adequate	over 35k	unknown	low risk
7	low	none	0-15k	bad	high-risk
8	low	adequate	over 35k	bad	Moderate risk
9	low	none	over 35k	good	
10	high	adequate	over 35k	good	
11	high	none	0-15k	good	high-risk
12	high	none	15-35k	good	Moderate risk

3 Perform data preprocessing on following dataset.

-Analyses on the dataset
-preprocess on null dataset
-what is dirty data set
-how to handle continues values
-differentiate between continues and discrete/categorical values.

Day	Outlook	Temperature	Humidity	Wind	Play ball
D1	SUNNY	HOT	HIGH	WEAK	NO

D2	SUNNY	HOT	HIGH	STRONG	NO	
D3	OVERCAST	HOT	HIGH	WEAK	YES	
D4	RAIN	MILD	HIGH	WEAK	YES	
D5	RAIN	COOL	NORMAL	WEAK	YES	
D6	RAIN	COOL	NORMAL	STRONG	NO	
D7	OVERCAST	COOL	NORMAL	STRONG	YES	
D8		MILD	HIGH		NO	
D9		COOL	NORMAL		YES	
D10	RAIN	MILD	NORMAL	WEAK	YES	
D11	SUNNY		NORMAL	STRONG	YES	
D12	OVERCAST		HIGH	STRONG	YES	
D13	OVERCAST	HOT	NORMAL	WEAK	YES	
D14	RAIN	MILD	HIGH	STRONG	NO	

4	Compute the standard deviations for the following sets: Set1 = [0,8,12,20,10,25,27,38,57] Set2 =[8,9,11,12,22,33,44,55,66] Justify which set has a larger spread and compute covariance for that set only.
	Compute the standard deviations for the following sets: Set1 = [0,8,12,20,25,30,40,50] Set2 =[8,9,11,12,15,17,29,30] Justify which set has a larger spread and compute covariance for that set only.
5	Apply PCA concepts on given 2X2 matrix. A= $\begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix}$

	Apply PCA concepts on given 2X2 matrix. A= $\begin{bmatrix} 6 & 4 \\ 5 & 10 \end{bmatrix}$
6	In the example below, the x-axis represents age OF CAR, and the y-axis represents speed. We have registered the age and speed of 13 cars as they were passing a tollbooth. Predict the speed of **23 years old car**, using in a linear regression. X(age of car) = [5,7,8,7,2,17,2,9,4,11,12,9,6] y (speed)= [99,86,87,88,111,86,103,87,94,78,77,85,86].
7	In the example below, the x-axis represents age OF CAR, and the y-axis represents speed. We have registered the age and speed of 13 cars as they were passing a tollbooth. Predict the speed of **18 years old car**, using in a linear regression. X(age of car) = [5,7,8,7,2,17,2,9,4,11,12,9,6] y (speed)= [99,86,87,88,111,86,103,87,94,78,77,85,86].
8	In the example below, the x-axis represents year of experience, and the y-axis represents salary. We have registered the details of 10 employees. Predict the salary for 15 years' experience, using in a linear regression. X(yr of experience)=[1.1, 1.3, 1.5, 2.2, 3.9, 4, 5.1, 5.3, 8.2, 9, 10.5] Y(salary)=[39343, 46205, 39891, 55794, 66029, 83088, 113812, 105582, 112635, 122391]
9	In the example below, the x-axis represents year of experience, and the y-axis represents salary. We have registered the details of 10 employees. Predict the salary for 23 years' experience, using in a linear regression. X(yr of experience)=[1.1, 1.3, 1.5, 2.2, 3.9, 4, 5.1, 5.3, 8.2, 9, 10.5] Y(salary)=[39343, 46205, 39891, 55794, 66029, 83088, 113812, 105582, 112635, 122391]
10	In the example below, the x-axis represents market spend for advertisement, and the y-axis represents profit. New Start-up Company spend money for advertisement tentative 10 events are given in database. Predict the profit for 9 lakh spend on marketing using in a linear regression.

	x(Market spend)=[471784.1, 443898.53, 407934.54, 383199.62, 366168.42, 362861.36, 127716.82, 323876.68, 311613.29, 304981.62] y(profit)=[192261.83, 191792.06, 191050.39, 182901.99, 166187.94, 156991.12, 156122.51, 155752.6, 152211.77, 149759.96]
11	In the example below, the x-axis represents market spend for advertisement, and the y-axis represents profit. New Start-up Company spend money for advertisement tentative 10 events are given in database. Predict the profit for 13 lakh spend on marketing using in a linear regression. x(Market spend)=[471784.1, 443898.53, 407934.54, 383199.62, 366168.42, 362861.36, 127716.82, 323876.68, 311613.29, 304981.62] y(profit)=[192261.83, 191792.06, 191050.39, 182901.99, 166187.94, 156991.12, 156122.51, 155752.6, 152211.77, 149759.96]
12	Design and implement MP model for AND GATE -Design Truth Table for AND GATE -Design NN network -Design ACTIVATION function -
13	Design and implement MP model for ANDNOT GATE -Design Truth Table for ANDNOT GATE -Design NN network -Design ACTIVATION function
14	Design and implement MP model for OR GATE -Design Truth Table for OR GATE -Design NN network

	-Design ACTIVATION function
15	Perform Navie Bayes algorithm using given dataset for unknown sample. **X= {OVERCAST, MILD, NORMAL,STRONG, ?}**
16	Perform Navie Bayes algorithm using given dataset for unknown sample.

Day	Outlook	Temperature	Humidity	Wind	Play ball
D1	SUNNY	HOT	HIGH	WEAK	NO
D2	SUNNY	HOT	HIGH	STRONG	NO
D3	OVERCAST	HOT	HIGH	WEAK	YES
D4	RAIN	MILD	HIGH	WEAK	YES
D5	RAIN	COOL	NORMAL	WEAK	YES
D6	RAIN	COOL	NORMAL	STRONG	NO
D7	OVERCAST	COOL	NORMAL	STRONG	YES
D8	SUNNY	MILD	HIGH	WEAK	NO
D9	SUNNY	COOL	NORMAL	WEAK	YES
D10	RAIN	MILD	NORMAL	WEAK	YES
D11	SUNNY	MILD	NORMAL	STRONG	YES
D12	OVERCAST	MILD	HIGH	STRONG	YES
D13	OVERCAST	HOT	NORMAL	WEAK	YES
D14	RAIN	MILD	HIGH	STRONG	NO

P{credit-History-bad, Debt- high, Collateral- none, Income-(15-35K)}

Sr.No	Debt	Collateral	Income	Credit-History	Risk
1	high	none	0-15k	bad	high-risk
2	high	none	15-35k	unknown	high-risk
3	low	none	15-35k	unknown	Moderate risk
4	low	none	0-15k	unknown	high-risk
5	low	none	over 35k	unknown	low risk
6	low	adequate	over 35k	unknown	low risk
7	low	none	0-15k	bad	high-risk
8	low	adequate	over 35k	bad	Moderate risk
9	low	none	over 35k	good	low risk
10	high	adequate	over 35k	good	low risk
11	high	none	0-15k	good	high-risk
12	high	none	15-35k	good	Moderate risk

17

Perform Navie Bayes algorithm using given dataset for unknown sample.

P(X|Cj)=p{Yes,no,female,yes,A}

Owns Home	Married	Gender	Employed	Creadit Rating	Risk Class
Yes	Yes	Male	Yes	A	B

No	No	Female	Yes	A	A
Yes	Yes	Female	Yes	B	C
Yes	No	Male	No	B	B
No	Yes	Female	Yes	B	C
No	No	Female	Yes	B	A
No	No	Male	No	B	B
Yes	No	Female	Yes	A	A
No	Yes	Female	Yes	A	C
Yes	Yes	Female	Yes	A	C

18 Apply K(=2)-Means algorithm over the data (185, 72), (170, 56), (168, 60), (179,68), (182,72), (188,77) up to two iterations and show the clusters. Initially choose first two objects as initial centroids.

19 The problem is as follows. You are given 15 points in the Cartesian coordinate system as follows. solve a numerical problem using k means clustering.

Point	Coordinates
A1	(2,10)
A2	(2,6)
A3	(11,11)

	A4	(6,9)	
	A5	(6,4)	
	A6	(1,2)	
	A7	(5,10)	
	A8	(4,9)	
	A9	(10,12)	
	A10	(7,5)	
	A11	(9,11)	
	A12	(4,6)	
	A13	(3,10)	
	A14	(3,8)	
	A15	(6,11)	
20	Use the k-means algorithm and Euclidean distance to cluster the following 8 examples into 3 clusters: A1=(2,10), A2=(2,5), A3=(8,4), A4=(5,8), A5=(7,5), A6=(6,4), A7=(1,2), A8=(4,9).		
21	Implement of k-means algorithm (using K=2)		

Individual	Variable 1	Variable 2
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

22 We have 4 medicines as our training data points object and each medicine has 2 attributes. Each attribute represents coordinate of the object. We have to determine which medicines belong to cluster 1 and which medicines belong to the other cluster.

Object	**Attribute1 (X): weight index**	**Attribute 2 (Y): pH**
Medicine A	**1**	**1**
Medicine B	**2**	**1**
Medicine C	**4**	**3**
Medicine D	**5**	**4**