

Subject : 2301CS503– Data Mining

Lab No	Part	Practical Definition																																
1	A	Introduction to Pandas library. Jupyter Notebook Exercise - 1																																
2	A	Perform the exercise using the function of panda library. Jupyter Notebook Exercise - 2																																
3	A	Perform Data Exploration Types of Attributes related tasks. Jupyter Notebook Exercise – 3																																
4	A	Perform Data Preprocessing of Chipotle dataset Jupyter Notebook Exercise – 4 Github exercise for advanced practice of panda’s function (Chipotle).																																
	B	Explore the orange widgets in orange tool application. (Structure)																																
5	A	Perform Data Preprocessing – Cleaning and Transformation Jupyter Notebook Exercise – 5 Apply Data Cleaning and Data Transformation on titanic Dataset.																																
	B	Explore the orange widgets in orange tool for above task.																																
6	A	Perform Data Preprocessing – Data Reduction (PCA). Jupyter Notebook Exercise – 6 Apply Data PCA iris flower dataset.																																
	B	Explore the orange widgets in orange tool for above task.																																
7	A	Apply apriori algorithm on given dataset, calculate support and confidence than find the association. <table><tr><th>TID</th><th>Items</th></tr><tr><td>100</td><td>1 3 4</td></tr><tr><td>200</td><td>2 3 5</td></tr><tr><td>300</td><td>1 2 3 5</td></tr><tr><td>400</td><td>2 5</td></tr></table> <table><tr><th>TID</th><th>Items</th></tr><tr><td>1</td><td>Bread, Milk</td></tr><tr><td>2</td><td>Bread, Diaper, Beer, Eggs</td></tr><tr><td>3</td><td>Milk, Diaper, Beer, Cola</td></tr><tr><td>4</td><td>Milk, Diaper, Beer, Cola</td></tr><tr><td>5</td><td>Bread, Milk, Diaper, Cola</td></tr></table> [Implement apriori for frequent patterns mining using java or python for lab no 5.]	TID	Items	100	1 3 4	200	2 3 5	300	1 2 3 5	400	2 5	TID	Items	1	Bread, Milk	2	Bread, Diaper, Beer, Eggs	3	Milk, Diaper, Beer, Cola	4	Milk, Diaper, Beer, Cola	5	Bread, Milk, Diaper, Cola										
TID	Items																																	
100	1 3 4																																	
200	2 3 5																																	
300	1 2 3 5																																	
400	2 5																																	
TID	Items																																	
1	Bread, Milk																																	
2	Bread, Diaper, Beer, Eggs																																	
3	Milk, Diaper, Beer, Cola																																	
4	Milk, Diaper, Beer, Cola																																	
5	Bread, Milk, Diaper, Cola																																	
8	A	Generate the Frequent Pattern Tree for the given dataset. <table><tr><th>TID</th><th>Items</th></tr><tr><td>1</td><td>EKMNOY</td></tr><tr><td>2</td><td>DEKNOY</td></tr><tr><td>3</td><td>AEKM</td></tr><tr><td>4</td><td>CKMU Y</td></tr><tr><td>5</td><td>CEIKO</td></tr></table> <table><tr><th>TID</th><th>Items</th></tr><tr><td>1</td><td>1 2 5</td></tr><tr><td>2</td><td>2 4</td></tr><tr><td>3</td><td>2 3</td></tr><tr><td>4</td><td>1 2 4</td></tr><tr><td>5</td><td>1 3</td></tr><tr><td>6</td><td>2 3</td></tr><tr><td>7</td><td>1 3</td></tr><tr><td>8</td><td>1 2 3 5</td></tr><tr><td>9</td><td>1 2 3</td></tr></table>	TID	Items	1	EKMNOY	2	DEKNOY	3	AEKM	4	CKMU Y	5	CEIKO	TID	Items	1	1 2 5	2	2 4	3	2 3	4	1 2 4	5	1 3	6	2 3	7	1 3	8	1 2 3 5	9	1 2 3
TID	Items																																	
1	EKMNOY																																	
2	DEKNOY																																	
3	AEKM																																	
4	CKMU Y																																	
5	CEIKO																																	
TID	Items																																	
1	1 2 5																																	
2	2 4																																	
3	2 3																																	
4	1 2 4																																	
5	1 3																																	
6	2 3																																	
7	1 3																																	
8	1 2 3 5																																	
9	1 2 3																																	
	B	Use orange tool to generate frequent pattern and association on second dataset having 9 transactions.																																

Subject : 2301CS503– Data Mining

9	A	<p>Apply Decision Tree for the given dataset.</p> <table><tr><th>Instance</th><th>a1</th><th>a2</th><th>a3</th><th>Classification</th></tr><tr><td>1</td><td>TRUE</td><td>HOT</td><td>HIGH</td><td>NO</td></tr><tr><td>2</td><td>TRUE</td><td>HOT</td><td>HIGH</td><td>NO</td></tr><tr><td>3</td><td>FALSE</td><td>HOT</td><td>HIGH</td><td>YES</td></tr><tr><td>4</td><td>FALSE</td><td>COOL</td><td>NORMAL</td><td>YES</td></tr><tr><td>5</td><td>FALSE</td><td>COOL</td><td>NORMAL</td><td>YES</td></tr><tr><td>6</td><td>TRUE</td><td>COOL</td><td>HIGH</td><td>NO</td></tr><tr><td>7</td><td>TRUE</td><td>HOT</td><td>HIGH</td><td>NO</td></tr><tr><td>8</td><td>TRUE</td><td>HOT</td><td>NORMAL</td><td>YES</td></tr><tr><td>9</td><td>FALSE</td><td>COOL</td><td>NORMAL</td><td>YES</td></tr><tr><td>10</td><td>FALSE</td><td>COOL</td><td>NORMAL</td><td>YES</td></tr></table>	Instance	a1	a2	a3	Classification	1	TRUE	HOT	HIGH	NO	2	TRUE	HOT	HIGH	NO	3	FALSE	HOT	HIGH	YES	4	FALSE	COOL	NORMAL	YES	5	FALSE	COOL	NORMAL	YES	6	TRUE	COOL	HIGH	NO	7	TRUE	HOT	HIGH	NO	8	TRUE	HOT	NORMAL	YES	9	FALSE	COOL	NORMAL	YES	10	FALSE	COOL	NORMAL	YES																																			
Instance	a1	a2	a3	Classification																																																																																								
1	TRUE	HOT	HIGH	NO																																																																																								
2	TRUE	HOT	HIGH	NO																																																																																								
3	FALSE	HOT	HIGH	YES																																																																																								
4	FALSE	COOL	NORMAL	YES																																																																																								
5	FALSE	COOL	NORMAL	YES																																																																																								
6	TRUE	COOL	HIGH	NO																																																																																								
7	TRUE	HOT	HIGH	NO																																																																																								
8	TRUE	HOT	NORMAL	YES																																																																																								
9	FALSE	COOL	NORMAL	YES																																																																																								
10	FALSE	COOL	NORMAL	YES																																																																																								
	B	<p>Perform Decision Tree using Orange Tool on above dataset.</p>																																																																																										
10	A	<p>[Implement Decision Tree using java or python for lab no 8.]</p>																																																																																										
11	A	<p>Apply Naïve Bayes Classifiers for the given dataset.</p> <table><tr><th>Day</th><th>Outlook</th><th>Temperature</th><th>Humidity</th><th>Wind</th><th>Play Tennis</th></tr><tr><td>1</td><td>Sunny</td><td>Hot</td><td>High</td><td>Weak</td><td>No</td></tr><tr><td>2</td><td>Sunny</td><td>Hot</td><td>High</td><td>Strong</td><td>No</td></tr><tr><td>3</td><td>Overcast</td><td>Hot</td><td>High</td><td>Weak</td><td>Yes</td></tr><tr><td>4</td><td>Rain</td><td>Mild</td><td>High</td><td>Weak</td><td>Yes</td></tr><tr><td>5</td><td>Rain</td><td>Cool</td><td>Normal</td><td>Weak</td><td>Yes</td></tr><tr><td>6</td><td>Rain</td><td>Cool</td><td>Normal</td><td>Strong</td><td>No</td></tr><tr><td>7</td><td>Overcast</td><td>Cool</td><td>Normal</td><td>Strong</td><td>Yes</td></tr><tr><td>8</td><td>Sunny</td><td>Mild</td><td>High</td><td>Weak</td><td>No</td></tr><tr><td>9</td><td>Sunny</td><td>Cool</td><td>Normal</td><td>Weak</td><td>Yes</td></tr><tr><td>10</td><td>Rain</td><td>Mild</td><td>Normal</td><td>Weak</td><td>Yes</td></tr><tr><td>11</td><td>Sunny</td><td>Mild</td><td>Normal</td><td>Strong</td><td>Yes</td></tr><tr><td>12</td><td>Overcast</td><td>Mild</td><td>High</td><td>Strong</td><td>Yes</td></tr><tr><td>13</td><td>Overcast</td><td>Hot</td><td>Normal</td><td>Weak</td><td>Yes</td></tr><tr><td>14</td><td>Rain</td><td>Mild</td><td>High</td><td>Strong</td><td>No</td></tr></table> <p>$x' = (\text{Outlook}=\text{Sunny}, \text{Temperature}=\text{Cool}, \text{Humidity}=\text{High}, \text{Wind}=\text{Strong})?$ is it Yes, or No? find it using naïve bayes classifiers.</p> <p>Apply Gini Index for the given dataset.</p>	Day	Outlook	Temperature	Humidity	Wind	Play Tennis	1	Sunny	Hot	High	Weak	No	2	Sunny	Hot	High	Strong	No	3	Overcast	Hot	High	Weak	Yes	4	Rain	Mild	High	Weak	Yes	5	Rain	Cool	Normal	Weak	Yes	6	Rain	Cool	Normal	Strong	No	7	Overcast	Cool	Normal	Strong	Yes	8	Sunny	Mild	High	Weak	No	9	Sunny	Cool	Normal	Weak	Yes	10	Rain	Mild	Normal	Weak	Yes	11	Sunny	Mild	Normal	Strong	Yes	12	Overcast	Mild	High	Strong	Yes	13	Overcast	Hot	Normal	Weak	Yes	14	Rain	Mild	High	Strong	No
Day	Outlook	Temperature	Humidity	Wind	Play Tennis																																																																																							
1	Sunny	Hot	High	Weak	No																																																																																							
2	Sunny	Hot	High	Strong	No																																																																																							
3	Overcast	Hot	High	Weak	Yes																																																																																							
4	Rain	Mild	High	Weak	Yes																																																																																							
5	Rain	Cool	Normal	Weak	Yes																																																																																							
6	Rain	Cool	Normal	Strong	No																																																																																							
7	Overcast	Cool	Normal	Strong	Yes																																																																																							
8	Sunny	Mild	High	Weak	No																																																																																							
9	Sunny	Cool	Normal	Weak	Yes																																																																																							
10	Rain	Mild	Normal	Weak	Yes																																																																																							
11	Sunny	Mild	Normal	Strong	Yes																																																																																							
12	Overcast	Mild	High	Strong	Yes																																																																																							
13	Overcast	Hot	Normal	Weak	Yes																																																																																							
14	Rain	Mild	High	Strong	No																																																																																							
	B	<table><tr><th>RID</th><th>Age</th><th>Income</th><th>Student</th><th>Credit_rating</th><th>Buys Computer</th></tr><tr><td>1</td><td>youth</td><td>high</td><td>no</td><td>fair</td><td>no</td></tr><tr><td>2</td><td>youth</td><td>high</td><td>no</td><td>excellent</td><td>no</td></tr><tr><td>3</td><td>middle_aged</td><td>high</td><td>no</td><td>fair</td><td>yes</td></tr><tr><td>4</td><td>senior</td><td>medium</td><td>no</td><td>fair</td><td>yes</td></tr><tr><td>5</td><td>senior</td><td>low</td><td>yes</td><td>fair</td><td>yes</td></tr><tr><td>6</td><td>Senior</td><td>low</td><td>yes</td><td>excellent</td><td>no</td></tr></table>	RID	Age	Income	Student	Credit_rating	Buys Computer	1	youth	high	no	fair	no	2	youth	high	no	excellent	no	3	middle_aged	high	no	fair	yes	4	senior	medium	no	fair	yes	5	senior	low	yes	fair	yes	6	Senior	low	yes	excellent	no																																																
RID	Age	Income	Student	Credit_rating	Buys Computer																																																																																							
1	youth	high	no	fair	no																																																																																							
2	youth	high	no	excellent	no																																																																																							
3	middle_aged	high	no	fair	yes																																																																																							
4	senior	medium	no	fair	yes																																																																																							
5	senior	low	yes	fair	yes																																																																																							
6	Senior	low	yes	excellent	no																																																																																							

Subject : 2301CS503– Data Mining

		<table><tr><td>7</td><td>middle_aged</td><td>low</td><td>yes</td><td>excellent</td><td>yes</td></tr><tr><td>8</td><td>youth</td><td>medium</td><td>no</td><td>fair</td><td>no</td></tr><tr><td>9</td><td>youth</td><td>low</td><td>yes</td><td>fair</td><td>yes</td></tr><tr><td>10</td><td>senior</td><td>medium</td><td>yes</td><td>fair</td><td>yes</td></tr><tr><td>11</td><td>youth</td><td>medium</td><td>yes</td><td>excellent</td><td>yes</td></tr><tr><td>12</td><td>middle_aged</td><td>medium</td><td>no</td><td>excellent</td><td>yes</td></tr><tr><td>13</td><td>middle_aged</td><td>high</td><td>yes</td><td>fair</td><td>yes</td></tr><tr><td>14</td><td>senior</td><td>medium</td><td>no</td><td>excellent</td><td>no</td></tr></table>	7	middle_aged	low	yes	excellent	yes	8	youth	medium	no	fair	no	9	youth	low	yes	fair	yes	10	senior	medium	yes	fair	yes	11	youth	medium	yes	excellent	yes	12	middle_aged	medium	no	excellent	yes	13	middle_aged	high	yes	fair	yes	14	senior	medium	no	excellent	no
7	middle_aged	low	yes	excellent	yes																																													
8	youth	medium	no	fair	no																																													
9	youth	low	yes	fair	yes																																													
10	senior	medium	yes	fair	yes																																													
11	youth	medium	yes	excellent	yes																																													
12	middle_aged	medium	no	excellent	yes																																													
13	middle_aged	high	yes	fair	yes																																													
14	senior	medium	no	excellent	no																																													
12	A	<p>Apply the k-mean clustering on the following data and form the cluster for the value of k=2.</p> <table><tr><th>Sr.</th><th>X</th><th>Y</th></tr><tr><td>1</td><td>1</td><td>1</td></tr><tr><td>2</td><td>1.5</td><td>2</td></tr><tr><td>3</td><td>3</td><td>4</td></tr><tr><td>4</td><td>5</td><td>7</td></tr><tr><td>5</td><td>3.5</td><td>5</td></tr><tr><td>6</td><td>4.5</td><td>5</td></tr><tr><td>7</td><td>3.5</td><td>4.5</td></tr></table>	Sr.	X	Y	1	1	1	2	1.5	2	3	3	4	4	5	7	5	3.5	5	6	4.5	5	7	3.5	4.5																								
Sr.	X	Y																																																
1	1	1																																																
2	1.5	2																																																
3	3	4																																																
4	5	7																																																
5	3.5	5																																																
6	4.5	5																																																
7	3.5	4.5																																																
	B	<p>Perform K-means cluster using Orange Tool on above dataset.</p>																																																
13	A	<p>Apply the k-mediods clustering on the following data and form the cluster for the value of k=2.</p> <table><tr><th>Sr.</th><th>X</th><th>Y</th></tr><tr><td>0</td><td>8</td><td>7</td></tr><tr><td>1</td><td>3</td><td>7</td></tr><tr><td>2</td><td>4</td><td>9</td></tr><tr><td>3</td><td>9</td><td>6</td></tr><tr><td>4</td><td>8</td><td>5</td></tr><tr><td>5</td><td>5</td><td>8</td></tr><tr><td>6</td><td>7</td><td>3</td></tr><tr><td>7</td><td>8</td><td>4</td></tr><tr><td>8</td><td>7</td><td>5</td></tr><tr><td>9</td><td>4</td><td>5</td></tr></table>	Sr.	X	Y	0	8	7	1	3	7	2	4	9	3	9	6	4	8	5	5	5	8	6	7	3	7	8	4	8	7	5	9	4	5															
Sr.	X	Y																																																
0	8	7																																																
1	3	7																																																
2	4	9																																																
3	9	6																																																
4	8	5																																																
5	5	8																																																
6	7	3																																																
7	8	4																																																
8	7	5																																																
9	4	5																																																
	B	<p>Perform K- mediods cluster using Orange Tool on above dataset.</p>																																																
14	A	<p>[Implement k-mean using java or python for lab no 12.]</p>																																																
	B	<p>[Implement k- Mediods using java or python for lab no 13.]</p>																																																

Subject : 2301CS503– Data Mining

15	A	<p>Apply the agglomerative clustering on the following data and form the cluster.</p> <table><tr><td></td><td>P1</td><td>P2</td><td>P3</td><td>P4</td><td>P5</td></tr><tr><td>P1</td><td>0</td><td></td><td></td><td></td><td></td></tr><tr><td>P2</td><td>9</td><td>0</td><td></td><td></td><td></td></tr><tr><td>P3</td><td>3</td><td>7</td><td>0</td><td></td><td></td></tr><tr><td>P4</td><td>6</td><td>5</td><td>9</td><td>0</td><td></td></tr><tr><td>P5</td><td>11</td><td>10</td><td>2</td><td>8</td><td>0</td></tr></table>		P1	P2	P3	P4	P5	P1	0					P2	9	0				P3	3	7	0			P4	6	5	9	0		P5	11	10	2	8	0
	P1	P2	P3	P4	P5																																	
P1	0																																					
P2	9	0																																				
P3	3	7	0																																			
P4	6	5	9	0																																		
P5	11	10	2	8	0																																	
	B	<p>Implement the same problem with its solution using python or java programming.</p>																																				
	C	<p>Apply the agglomerative clustering on the following data and form the cluster.</p> <p>https://www.kaggle.com/datasets/rajasankar/hierarchical-clustering-of-7-million-proteins</p>																																				

Subject : 2301CS503– Data Mining