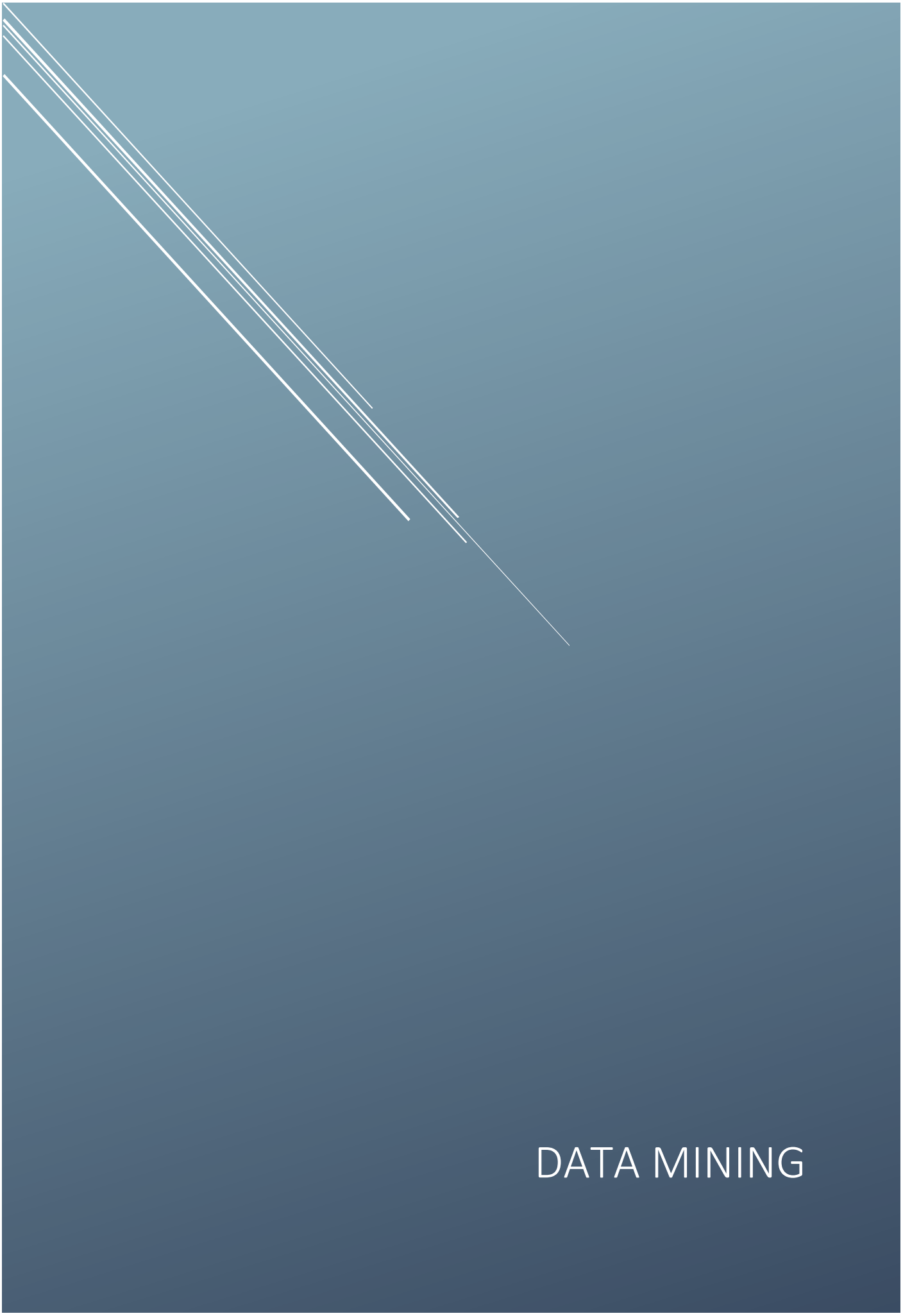
DATA MINING

Data Mining Assignments

اسم الطالب /

عبدالرحمن عبدالجليل عبدالرحمن

اشراف الدكتورة /

هبة المروعي

Association Rules(Eclat)

Introduction

Eclat algorithm stands for Equivalence Class Transformation**.** It's a bottom up lattice traversal. It is more efficient and scalable version of Apriori algorithm .

While the Apriori algorithm works in a horizontal sense imitating the Breadth-First Search of a graph, the ECLAT algorithm works in a vertical manner just like the Depth-First Search of a graph. This vertical approach of the ECLAT algorithm makes it a faster algorithm than the Apriori algorithm.

Consider the following transactions record

Transaction Id	Bread	Butter	Milk	Coke	Jam
T1	1	1	0	0	1
T2	0	1	0	1	0
T3	0	1	1	0	0
T4	1	1	0	1	0
T5	1	0	1	0	0
T6	0	1	1	0	0
T7	1	0	1	0	0
T8	1	1	1	0	1
T9	1	1	1	0	0

The above-given data is a boolean matrix where for each cell (i, j), the value denotes whether the j'th item is included in the i'th transaction or not. 1 means true while 0 means false.

We now call the function for the first time and arrange each item with it's tidset in a tabular fashion:-

Confidence = 70% , minimum support = 2

Solution:

Step 1: Convert DB into vertical format with min_support_cnt gets satisfied K = 1

Item_Set	List of Items
I1 (Bread)	T1,T4,T5,T7,T8,T9
I2 (Butter)	T1,T2,T3,T4,T6,T8,T9
I3 (Milk)	T3,T5,T6,T7,T8,T9
I4 (Coke)	T2,T4
I5 (Jam)	T1,T8

Step 2: @k = 2

Item_Set	List of Items
I1,I2	T1,T4,T8,T9
I1,I3	T5,T7,T8,T9
I1,I4	T4
I1,I5	T1,T8
I2,I4	T2,T4
I2,I5	T1,T8
I3,I4	NULL
I3,I5	T8
I4,I5	NULL

Eliminate the Item_set's with less than 2 list of items

Item_Set	List of Items
I1,I2	T1,T4,T8,T9
I1,I3	T5,T7,T8,T9
I1,I5	T1,T8
I2,I4	T2,T4
I2,I5	T1,T8

Step 3: @k = 3

Item_Set	List of Items
I1,I2,I3	T8,T9
I1,I2,I5	T1,T8
I1,I3,I5	T8
I2,I3,I4	NULL
I2,I3,I5	T8
I2,I4,I5	NULL

Eliminate the Item_set's with less than 2 list of items

Item_Set	List of Items
I1,I2,I3	T8,T9
I1,I2,I5	T1,T8

Step:4 @k = 4

Item_Set	List of Items
I1,I2,I3,I5	T8

Here minimum confidence count does not match, so go to previous step.

@step:3 we have two itemsets and they are

{1,2,3} => {1,2},{1,3},{2,3},{1},{2},{3} => equation — 1

{1,2,5} => {1,2},{1,5},{2,5},{1},{2},{5} => equation — 2

Applying rules on equation — 1

Rule 1 : {1,2} => {1,2,3} and {1,2}.

Confidence = (1,2,3) / (1,2) => 2/4 => 0.5

Rule 2: {2,3} => {1,2,3} and {2,3}.

Confidence = (1,2,3) / (2,3) => 2/4 => 0.5

Rule 3 : {1,3} => {1,2,3} and {1,3}.

Confidence = (1,2,3) / (1,3) => 2/4 => 0.5

Rule 4: {1} => {1,2,3} and {1}.

Confidence = (1,2,3) / (1) => 2/6 => 0.33

Rule 5: {2} => {1,2,3} and {2}.

Confidence = (1,2,3) / (2) => 2/7 => 0.29

Rule 6: {3} => {1,2,3} and {3}.

Confidence = (1,2,3) / (3) => 2/6 => 0.33

None of the combination satisfies at equation 1

Applying rules on equation — 2

Rule 1 : {1,2} => {1,2,5} and {1,2}.

Confidence = (1,2,5) / (1,2) => 2/4 => 0.5

Rule 2: {2,5} => {1,2,5} and {2,5}.

Confidence = (1,2,5) / (2,5) => 2/2 => 1.0

Rule 3 : {1,5} => {1,2,5} and {1,5}.

Confidence = (1,2,5) / (1,5) => 2/2 => 1.0

Rule 4: {1} => {1,2,5} and {1}.

Confidence = (1,2,5) / (1) => 2/6 => 0.33

Rule 5: {2} => {1,2,5} and {2}.

Confidence = (1,2,5) / (2) => 2/7 => 0.29

Rule 6: {5} => {1,2,5} and {5}.

Confidence = (1,2,5) / (5) => 2/2 => 1.00

Only Rule of the combinations 2,3 and 6 satisfies at equation 2

From overall result most commonly bought products with high confidence(>70%) are

Bread, Butter and Jam

Bread, Butter and Coke

Advantages over Apriori algorithm:-

1. **Memory Requirements:** Since the ECLAT algorithm uses a Depth-First Search approach, it uses less memory than Apriori algorithm.
2. **Speed:** The ECLAT algorithm is typically faster than the Apriori algorithm.
3. **Number of Computations:** The ECLAT algorithm does not involve the repeated scanning of the data to compute the individual support values.