SYMBOLIC DATA MINING

- **Frequent**: If $supp(X) \geq min\text{-}supp$, then X is frequent
- **Closed-Itemset**: An itemset, doesn't have a superset with same supp
- **Generator**: An itemset, doesn't have a subset with same supp

	A	B	C	D	E
1	X	X		X	X
2	X		X		
3	X	X	X		X
4		X	X		X
5	X	X	X		X

Itemset: A, AB, BCE...

Support(A)=4

Support(AB)=3

Support(BCE)=3

Min-Support: Threshold

One Size Larger Superset:

| SAME | X |

+

| SAME | Y |

⇓

| SAME | X | Y |

UNIVERSITY OF DEBRECEN
Kerem Dizen, PhD

APRIORI (breadth-first)(levelwise)

	A	B	C	D	E
1	X	X		X	X
2	X		X		
3	X	X	X		X
4		X	X		X
5	X	X	X		X

min-supp = 2

FC_1

A	4
B	4
C	4
~~D~~	~~1~~
E	4

→ F_1

A	4
B	4
C	4
E	4

FC_2

AB	3
AC	3
AE	3
BC	3
BE	4
CE	3

→ F_2

AB	3
AC	3
AE	3
BC	3
BE	4
CE	3

FC_3

ABC	2
ABE	3
ACE	2
BCE	3

→ F_3

ABC	2
ABE	3
ACE	2
BCE	3

FC_4

ABCE	2

→ F_4

ABCE	2

FC_5

∅ stop!

$|F_5| = 15$

UNIVERSITY OF DEBRECEN
Kerem Düzenli PhD

APRIORI-CLOSE (breadth-first) (levelwise)

	A	B	C	D	E
1	X	X		X	X
2	X		X		
3	X	X	X		X
4		X	X		X
5	X	X	X		X

min-supp = 3

FC_1

A	4
B	4
C	4
D	1
E	4

F_1

itemset	support	closed?
A	4	yes
B	4	~~yes~~ no
C	4	yes
E	4	~~yes~~ no

FC_2

A B	3
A C	3
A E	3
B C	3
B E	4
C E	3

F_2

itemset	support	closed?
AB	3	~~yes~~ no
AC	3	yes
AE	3	~~yes~~ no
BC	3	~~yes~~ no
BE	4	yes
CE	3	~~yes~~ no

FC_3

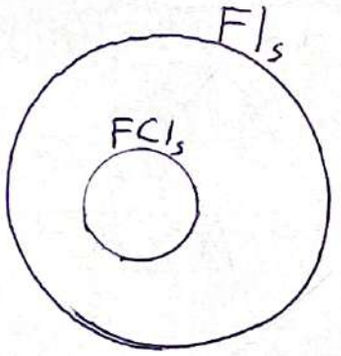
~~ABC~~	~~2~~
ABE	3
~~ACE~~	~~2~~
BCE	3

F_3

itemset	support	closed?
ABE	3	yes
BCE	3	yes

FC_4

$\emptyset$ stop!



$FI_S = 12$

$FCI_S = 6$

UNIVERSITY OF DEBRECEN
Kerem Düzenli, PhD

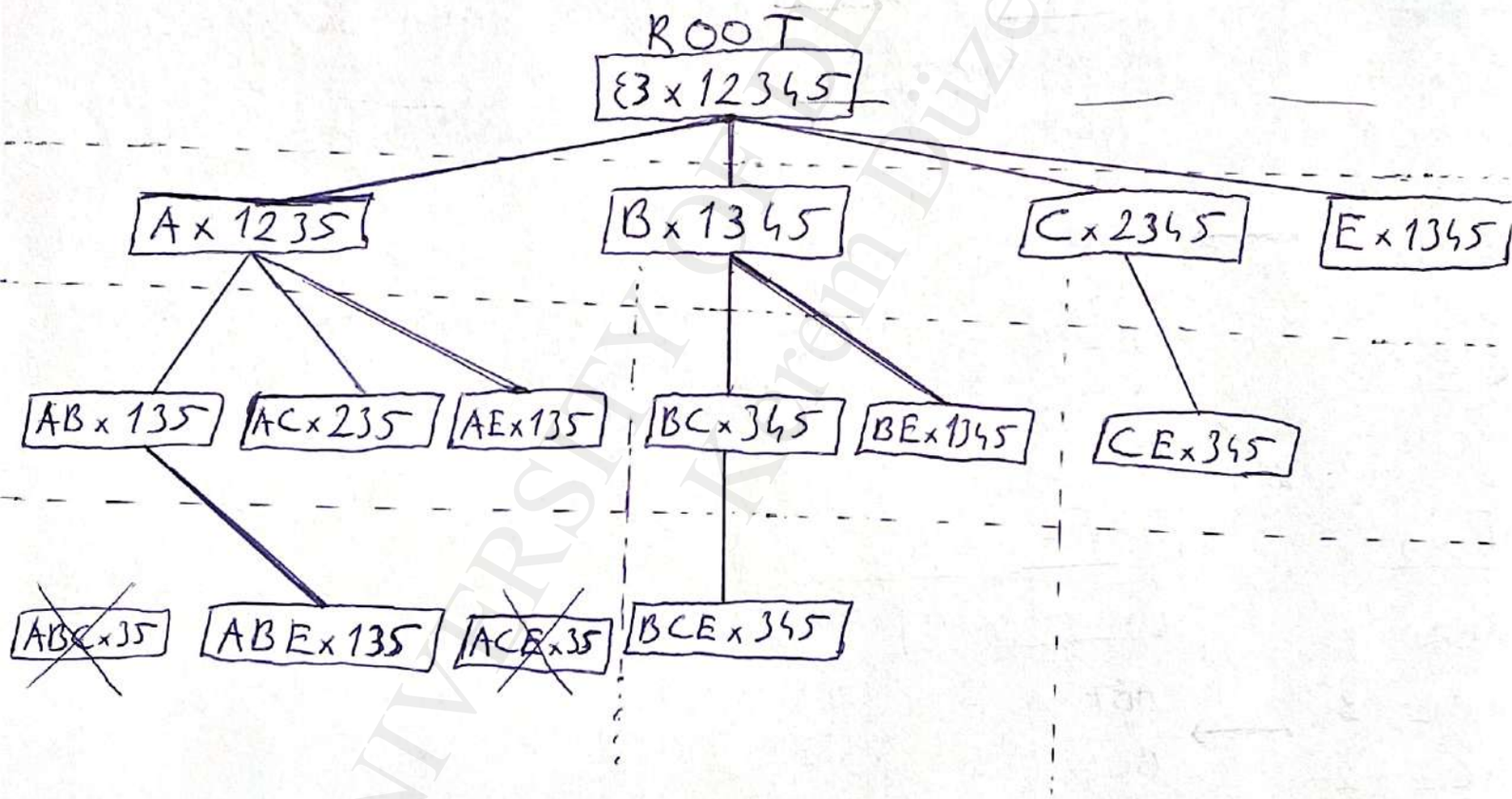
ECLAT (vertical) (depth-first)

1 → ABDE
2 → AC
3 → ABCE
4 → BCE
5 → ABCE

A	B	C	D	E
1	1	2	1	1
2	3	3		3
3	4	4		4
5	5	5		5

	A	B	C	D	E
1	X	X		X	X
2	X		X		
3	X	X	X		X
4		X	X		X
5	X	X	X		X

min-supp=3



UNIVERSITY OF DEBRECEN
Kerem Düzenli, PhD

ASSOCIATION RULE

	A	B	C	D	E
1	X	X		X	X
2	X		X		
3	X	X	X		X
4		X	X		X
5	X	X	X		X

min-supp = 3

FC_1

A	4
B	4
C	4
D	1
E	4

F_1

A	4
B	4
C	4
E	4

FC_2

AB	3
AC	3
AE	3
BC	3
BE	4
CE	3

F_2

AB	3
AC	3
AE	3
BC	3
BE	4
CE	3

FC_3

ABC	2
ABE	3
ACE	2
BCE	3

F_3

ABE	3
BCE	3

FC_4

$\emptyset$

$Fl_3 = 12$

2 Steps:

1) Find all Fl_s
2) Generate association rules from Fl_s

$r: A \rightarrow B$

$supp(r) = supp(A \cup B) \longrightarrow 3$

$conf(r) = \dfrac{supp(A \cup B) \longrightarrow 3}{supp(A) \longrightarrow 4} = \%75$

$r: BE \rightarrow A$

$supp(r) = supp(ABE) \longrightarrow 3$

$conf(r) = \dfrac{supp(ABE) \longrightarrow 3}{supp(BE) \longrightarrow 4} = \%75$

UNIVERSITY OF DEBRECEN
Keren Büzenli PhD

APRIORI - RARE

	A	B	C	D	E
1	X	X		X	X
2	X		X		
3	X	X	X		X
4		X	X		X
5	X	X	X		X

min-supp = 3

FC_1

A 4
B 4
C 4
D 1
E 4

R_1

D 1

F_1

itemset	support	closed
A	4	yes
B	4	~~yes~~ no
C	4	yes
E	4	~~yes~~ no

FC_2

AB 3
AC 3
AE 3
BC 3
BE 4
CE 4

R_2

∅

F_2

itemset	support	closed
AB	3	~~yes~~ no
AC	3	yes
AE	3	~~yes~~ no
BC	3	~~yes~~ no
BE	4	yes
CE	3	~~yes~~ no

FC_3

ABC 2
ABE 3
ACE 2
BCE 3

R_3

ABC 2
ACE 2

F_3

itemset	support	closed
ABE	3	yes
BCE	3	yes

FC_4

∅

$FI_s = 12$

$FCI_s = 6$

Minimal-Rare = 3

- **Minimal-Rare Itemset**: An itemset that is not frequent, but all subsets are frequent

UNIVERSITY OF DEBRECEN
Kerem Düzenli PhD