# OBJECTIVE

**Mining of structured data to find potentially useful patterns by Association Rule Mining.**

**Traditional Algorithms Already Available like Apriori, FP-Growth etc.**

# Why new approach?

1. No need to search entire lattice of item combinations.

2. No pruning step required.

3. Less computational resources in terms of time and memory.

4. k-items patterns can be obtained easily.

5. The proposed algorithm generate tree which is useful in visual analysis of data.

# What is the task?

: **Structured data**

| A1 | A2 | A3 | A4 | A5 |
|-----|-----|-----|-----|-----|
| V11 | V21 | V31 | V41 | V51 |
| V11 | V22 | V32 | V42 | V51 |

**Where Ai are the column names? and Vij are the values in the ith column.**

**INPUT 2** : **Threshold support and threshold confidence**

**OUTPUT** : **Rules satisfying threshold support and threshold confidence.**

**For example : (A1 = V11 => A5 = V51)**

# Rule generation will be done in two steps

**Step 1** : **Generate MASP Tree**

**Step 2** : **Generate rules using MASP Tree**

# MASP tree generation

| | A1 | A2 | A3 | A4 | A5 |
|----|----|----|----|----|----|
| 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 2 | 1 | 3 |
| 3 | 1 | 2 | 1 | 2 | 2 |
| 4 | 1 | 1 | 2 | 2 | 2 |
| 5 | 1 | 2 | 2 | 3 | 2 |
| 6 | 1 | 2 | 2 | 3 | 2 |
| 7 | 2 | 2 | 1 | 4 | 1 |
| 8 | 2 | 2 | 1 | 4 | 1 |
| 9 | 2 | 3 | 1 | 5 | 1 |
| 10 | 2 | 3 | 2 | 6 | 1 |

**Threshold support** = 0.2
**Threshold confidence** = 0.3

# MASP tree generation

## Step 1: Generate item distribution of the candidate block

| Item | Frequency | Item | Frequency |
|------|-----------|------|-----------|
| A1 = 1 | 6 | A4 = 2 | 2 |
| A1 = 2 | 4 | A4 = 3 | 2 |
| A2 = 1 | 3 | A4 = 4 | 2 |
| A2 = 2 | 5 | A4 = 5 | 1 |
| A2 = 3 | 2 | A4 = 6 | 1 |
| A3  = 1 | 5 | A5 = 1 | 5 |
| A3 = 2 | 5 | A5 = 2 | 4 |
| A4 = 1 | 2 | A5 = 3 | 1 |

# MASP tree generation

**Step 2: Select item having maximum frequency**

**Step 3: Check whether it is satisfying threshold support and Confidence**

**Step 4: If threshold is not satisfied stop the algorithm**

# MASP tree generation

In the above data  A1 = 1 frequency is 6 (maximum) and support = confidence = 6/10 = 0.6 (greater than the threshold)


Algorithm will not stop !

# MASP tree generation

## Split the data based on A1 = 1

|     | A1 | A2 | A3 | A4 | A5 |
| --- | --- | --- | --- | --- | --- |
| 1   | 1  | 1  | 1  | 1  | 1  |
| 2   | 1  | 1  | 2  | 1  | 3  |
| 3   | 1  | 2  | 1  | 2  | 2  |
| 4   | 1  | 1  | 2  | 2  | 2  |
| 5   | 1  | 2  | 2  | 3  | 2  |
| 6   | 1  | 2  | 2  | 3  | 2  |
| 7   | 2  | 2  | 1  | 4  | 1  |
| 8   | 2  | 2  | 1  | 4  | 1  |
| 9   | 2  | 3  | 1  | 5  | 1  |
| 10  | 2  | 3  | 2  | 6  | 1  |

# MASP tree generation

**Left part of the root will be A1 = 1 and Right part will be ~A1 = 1 and the data associated with**

**A1 = 1**

|   | A2 | A3 | A4 | A5 |
|---|----|----|----|----|
| 1 | 1  | 1  | 1  | 1  |
| 2 | 1  | 2  | 1  | 3  |
| 3 | 2  | 1  | 2  | 2  |
| 4 | 1  | 2  | 2  | 2  |
| 5 | 2  | 2  | 3  | 2  |
| 6 | 2  | 2  | 3  | 2  |

**~A1 = 1**

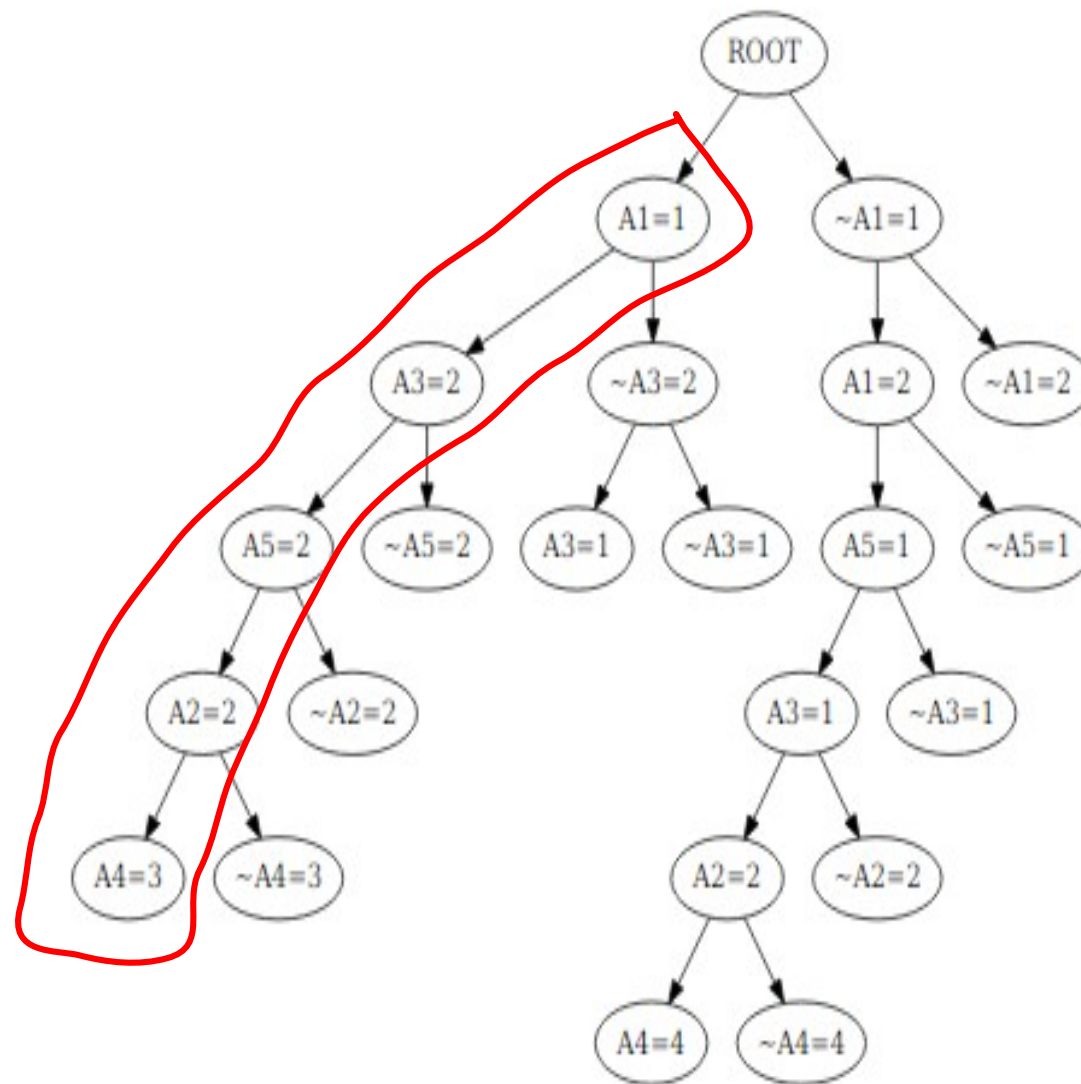|   | A1 | A2 | A3 | A4 | A5 |
|---|----|----|----|----|----|
| 1 | 2  | 2  | 1  | 4  | 1  |
| 2 | 2  | 2  | 1  | 4  | 1  |
| 3 | 2  | 3  | 1  | 5  | 1  |
| 4 | 2  | 3  | 2  | 6  | 1  |

# MASP tree generation

**Recursively apply the same steps on newly generated data**

# MASP tree generation
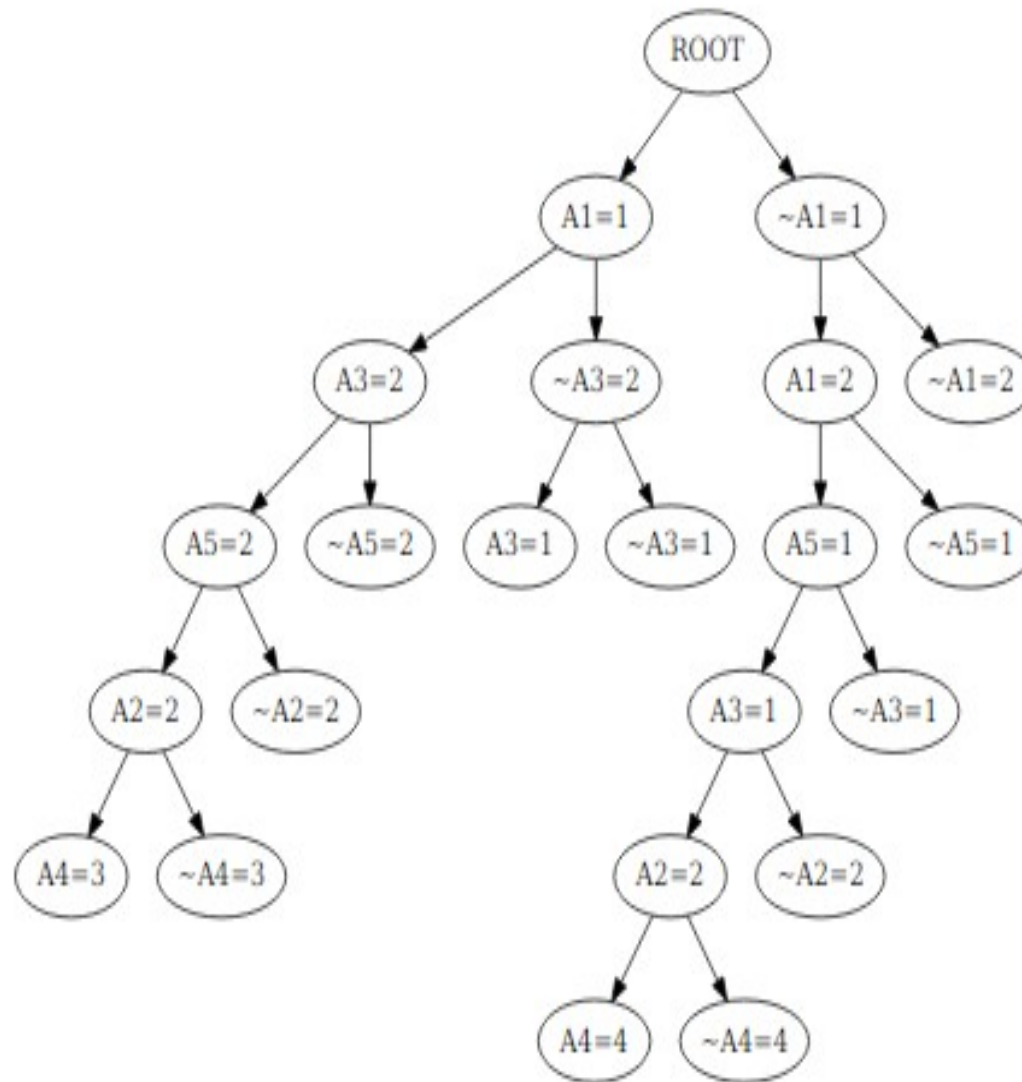
# Rules Generation

# Rules Generation

**Path selected**

(A1 = 1) -> (A3 = 2) -> (A5 = 2) -> (A2 = 2) -> (A4 = 3)

**Rules (easy to generate)**

(A1 = 1) (A3 = 2) (A5 = 2) (A2 = 2) => (A4 = 3)

(A1 = 1) (A3 = 2) (A5 = 2) => (A2 = 2)

(A1 = 1) (A3 = 2) => (A5 = 2)

(A1 = 1) => (A3 = 2)

# Thank You