# Association Rules

# What Are Association Rules?

- Rules of the form: `IF X THEN Y`
- `X` and `Y` are *itemsets* (collections of attribute–value pairs)
- Widely used in *market basket analysis*
- Goal: find *frequent co-occurring patterns*
- Example:

```
IF humidity = high AND windy = false -> play = no
```

# Why Association Rules?

- Reveal **hidden correlations**
- Useful in:
  - Retail analysis
  - Web usage mining
  - Medical diagnosis
  - Bioinformatics
- Not limited to classification — predicts **any attribute combination**

# Key Concepts

- **Item**
  - Single attribute=value pair
  - Example: `outlook = sunny`
- **Itemset**
  - A set of items
  - Example: `{outlook = rainy, windy = false}`
- **Support**
  - Frequency of an itemset in dataset
- **Confidence**
  - Accuracy of a rule given its antecedent

# Support & Confidence Definitions

- For rule $\mathbf{X} \rightarrow \mathbf{Y}$:
  - Support = `(# instances containing X and Y) / N`
  - Confidence = `Support(X union Y) / Support(X)`
- Example:

```
Support = 4/14
Confidence = (4/14) / (6/14) = 0.67
```

# Association Rule Process

1. Find frequent itemsets
2. Generate candidate rules
3. Compute support and confidence
4. Meets thresholds?
    - Yes: Add rule
    - No: Discard

# The Apriori Principle

- Any subset of a frequent itemset must also be frequent.
- This enables efficient pruning:
  - If `{A, B}` is infrequent -> ignore `{A, B, C}`
- This reduces combinatorial explosion.

# Frequent Itemset Mining Steps

1. Count all **1-itemsets**
2. Keep those meeting minimum support
3. Generate **candidate 2-itemsets**
4. Count occurrences
5. Keep those meeting support
6. Repeat for k-itemsets

Continue until no more itemsets meet minimum support.

# Example: Weather Data Itemsets

- Example 1-itemsets (support $>=$ 2):

```
outlook = sunny (5)
temperature = cool (4)
humidity = high (7)
play = yes (9)
```

- Example 2-itemsets:

```
outlook = sunny AND humidity = high (3)
temperature = cool AND windy = false (4)
```

# Rule Generation from Itemsets

- For `{A, B, C}` generate rules:

  `A, B -> C`
  `A, C -> B`
  `B, C -> A`
  `A → B, C`
  `B → A, C`
  `C → A, B`

- Each rule must meet:
  - Minimum support
  - Minimum confidence

# Example Rule Evaluation

- Itemset:

  `{humidity = normal, windy = false, play = yes}`

- Rule:

  `humidity = normal AND windy = false -> play = yes`

- Support $= 4$
- Confidence $= 4/4 = 100\%$
- This rule is accepted.

# Why So Many Rules?

- A single itemset may produce **many** rules.
- Example:

  `{temperature = cool, humidity = normal, windy = false,`

  - Can generate >= **4** rules.
- Only rules meeting confidence threshold survive.

# Avoiding Rule Explosion

- We prune rules via:
  - Minimum support
- Filters rare itemsets early
  - Minimum confidence
- Removes weak rules
- Optional filters:
  - Max rule length
  - Lift / conviction measures

# Algorithm Efficiency

- Apriori improves efficiency by:
  - Early pruning of candidate itemsets
  - Hash-based counting
  - Reducing dataset scans
- However, can still be expensive for:
  - Large attribute sets
  - Very low support thresholds
- Modern improvements:
  - FP-growth
  - Eclat

# Example High-Quality Rules

- Rules from weather data (support $>=$ 2, confidence $=$ 100%):

```
humidity = normal AND windy = false -> play = yes
outlook = overcast -> play = yes
temperature = cool -> humidity = normal
outlook = sunny AND humidity = high -> play = no
```

- These are highly interpretable.

# Strengths of Association Rules

- Finds *hidden patterns*
- Easy to interpret
- Works well on large datasets
- No target variable required
- Domain-independent

# Limitations

- Can produce large rule sets
- Many rules may be redundant or trivial
- Requires good support/confidence thresholds
- May need advanced measures for filtering

# Summary

- Association rules identify *frequent co-occurring patterns*
- Apriori mines frequent itemsets via *downward closure*
- Rules are filtered using:
  - Support
  - Confidence
- Effective across many data-driven applications
- Must control rule explosion