# Jaccard Coefficient Calculation

To calculate the Jaccard coefficient, we compare shared attributes between pairs of individuals.

Formula:

Jaccard Coefficient = (Number of Matching Attributes) / (Number of Compared Attributes)

We treat 'A' (Absent) as missing and skip those comparisons.

Data Table (Simplified):

| Name | Fever | Cough | Test-1 | Test-2 | Test-3 | Test-4 |
|---|---|---|---|---|---|---|
| Jack | Y | N | P | N | N | A |
| Mary | Y | N | P | A | P | N |
| Jim | Y | P | N | N | N | A |

Pair: (Jack, Mary)

Compared Attributes: Fever, Cough, Test-1, Test-3

Matches: Fever, Cough, Test-1 -> 3

Total: 4 -> Jaccard = 3 / 4 = 0.75

Pair: (Jack, Jim)

Compared Attributes: Fever, Cough, Test-1, Test-2, Test-3

Matches: Fever, Test-2, Test-3 -> 3

Total: 5 -> Jaccard = 3 / 5 = 0.6

Pair: (Jim, Mary)

Compared Attributes: Fever, Cough, Test-1, Test-3

Matches: Fever -> 1

Total: 4 -> Jaccard = 1 / 4 = 0.25

Final Results:

(Jack, Mary):  0.75

(Jack, Jim):  0.60

(Jim, Mary):  0.25

Learning Outcomes:

- Machine learning professionals must handle missing data (e.g., 'A') carefully.

- Understand dataset challenges: incompleteness, ambiguity, and data interpretation.

- Ensure ethical, legal, and social awareness when using health-related data.