

## PRACTICAL 7

<b>Name:</b>	Harsh Shah	<b>Semester:</b>	VII	<b>Division:</b>	6
<b>Roll No.:</b>	21BCP359	<b>Date:</b>	10-09-24	<b>Batch:</b>	G11
<b>Aim:</b>	Understanding Jaccard Similarity.				

### Jaccard Similarity

Jaccard Similarity is a measure of similarity between two asymmetric binary vectors or we can say a way to find the similarity between two sets. It is a common proximity measurement used to compute the similarity of two items, such as two text documents. The index ranges from 0 to 1. Range closer to 1 means more similarity in two sets of data.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

### Code

```
def jaccard_similarity(list1, list2):
```

```
    intersection = 0
    union = 0
```

```
    for a, b in zip(list1, list2):
        if a == 1 or b == 1:
            union += 1
        if a == 1 and b == 1:
            intersection += 1
```

```
    if union == 0:
        return 0
    return intersection / union
```

```
C1 = [0, 1, 0, 0, 0, 1, 0, 0, 1]
```

```
C2 = [0, 0, 1, 0, 0, 0, 0, 0, 1]
```

```
C3 = [1, 1, 0, 0, 0, 1, 0, 0, 0]
```

```
similarity_C1_C2 = jaccard_similarity(C1, C2)
```

```
similarity_C1_C3 = jaccard_similarity(C1, C3)
```

```
similarity_C2_C3 = jaccard_similarity(C2, C3)
```

```
print(f'Similarity - Customer C1 and C2 is {similarity_C1_C2}')
```

```
print(f'Similarity - Customer C1 and C3 is {similarity_C1_C3}')
```

```
print(f'Similarity - Customer C2 and C3 is {similarity_C2_C3}')
```

```
def jaccard_similarity_sets(set1, set2):
```

```
    intersection = len(set(set1).intersection(set2))
```

```
union = len(set(set1).union(set2))  
return intersection / union
```

```
S1 = [0, 2, 5, 7, 9]  
S2 = [0, 1, 2, 4, 5, 6, 8]
```

```
similarity_S1_S2 = jaccard_similarity_sets(S1, S2)
```

```
print(f'Similarity between Set S1 and S2 is {similarity_S1_S2}')
```

## Output

```
Similarity - Customer C1 and C2 is 0.25  
Similarity - Customer C1 and C3 is 0.5  
Similarity - Customer C2 and C3 is 0.0  
Similarity between Set S1 and S2 is 0.3333333333333333
```