- ✓ Term Frequency (TF)
- √ Inverse Document Frequency (IDF)
- ✓ TF-IDF Vectorizer

Dataset:

Sentence1 = Love Bangladesh

Sentence2 = Love Germany

Sentence3 = Love Bangladesh Germany

| Word | Frequency | |
|------------|-----------|--|
| Love | 3 | |
| Bangladesh | 2 | |
| Germany | 2 | |

Term Frequency (TF) = Number of repetition of word in sentence

Total word in sentence

| Word | Sentence1 | Sentence2 | Sentesnce3 |
|------------|-----------|-----------|------------|
| Love | 1/2 | 1/2 | 1/2 |
| Bangladesh | □ 1/2 | 0 | 1/3 |
| Germany | 0 | 1/2 | 1/3 |

Dataset:

Sentence1 = Love Bangladesh

Sentence2 = Love Germany

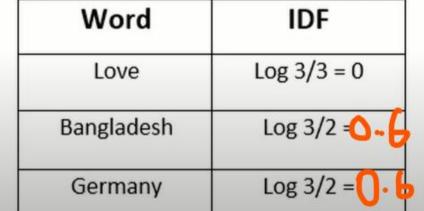
Sentence3 = Love Bangladesh Germany

| Word | Frequency | |
|------------|-----------|--|
| Love | 3 | |
| Bangladesh | 2 | |
| Germany | 2 | |

IDF = Total number of sentence

No of sentence contain in the word

B



| Word | Sentence1 | Sentence2 | Sentesnce3 |
|------------|-----------|-----------|------------|
| Love | 1/2 | 1/2 | 1/2 |
| Bangladesh | 1/2 | 0 | 1/3 |
| Germany | 0 | 1/2 | 1/3 |

| Word | IDF | |
|------------|-------------|--|
| Love | Log 3/3 = 0 | |
| Bangladesh | Log 3/2 = | |
| Germany | Log 3/2 = | |

B

TF * IDF =

| Sentences | Love | Bangladesh | Germany |
|-----------|------|---------------|---------------|
| Sentence1 | 0 | 1/2 * log 3/2 | 0 |
| Sentence2 | 0 | 0 | 1/2 * log 3/2 |
| Sentence3 | 0 | 1/3 * log 3/2 | 1/3* log 3/2 |