D 1: This is the example fot the TF-IDF section

D 2: The chatbot is good

*Step 1: Term Frequency: measures the number of times a word occurs in a document*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| D 1 | this | is | the | example | for | the | TF-IDF | section |
| Term Frequency | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 |

Table 3. TF for D 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| D 2 | the | chatbot | is | good |
| Term Frequency | 1 | 1 | 1 | 1 |

Table 4. TF for D 2

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| D 1 | this | is | the | example | for | the | TF-IDF | section |
| Normalized TF | 1/8 | 1/8 | 2/8 | 1/8 | 1/8 | 2/8 | 1/8 | 1/8 |

Table 6. Normalized TF for D 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| D 2 | the | chatbot | is | good |
| Term Frequency | 1/4 | 1/4 | 1/4 | 1/4 |

Table 7. Normalized TF for D 2

*Step 2: Inverse Document Frequency:*

IDF(the) = loge(Total Number Of Documents / Number Of Documents with term “the” in it) = loge(2/2)=0

If a word has appeared in all documents, then probably that word is not relevant to a particular document, but if it has appeared in a subset of documents then probably the word may be relevant to the documents it is presented in.

*Step 3: TF \* IDF*

→ vector for ‘the’ = (0, 0)