Compute the TF-IDF values for the terms in these three reviews/documents.

1. `['food', 'delici', 'servic', 'excel']`

2. `['oh', 'wow', 'place', 'amaz', '…', 'love', 'wait', 'hour', 'food']`

3. `['absolut', 'love', 'phone', 'batteri', 'last', 'day']`

Step 1: Compute Term Frequency (TF)

First, we calculate the Term Frequency (TF) for each term in each document.

- Document 1
  - Total terms: 4
  - TF: food: ¼ , delici: ¼,servic: ¼, excel: ¼
- Document 2
  - Total terms: 9
  - TF: oh: $1/9$,wow: $1/9$, place: $1/9$, amaz: $1/9$, …: $1/9$, love: $1/9$, wait: $1/9$, hour: $1/9$, food: $1/9$
- Document 3
  - Total terms: 6
  - TF: absolut: $1/6$,love: $1/6$,phone: $1/6$,batteri: $1/6$,last: $1/6$,day: $1/6$

Step 2: Compute Inverse Document Frequency (IDF)

Next, we calculate the Inverse Document Frequency (IDF) for each term.

$$IDF(t) = \log\left(\frac{N}{1+df(t)}\right)$$

Where:

- t is the term ("food" in this case)
- N is the total number of documents
- df(t) is the number of documents containing the term t

For the term "food":

- N = 3 (since there are 3 documents)
- df(food) = 2 (the term "food" appears in 2 documents

N = 3 (total number of documents)

Term	df(t)	IDF(t)
food	2	$\log\left(\frac{3}{1+2}\right) = 0.18$
delici	1	$\log\left(\frac{3}{1+1}\right) = 0.48$
servic	1	0.48
excel	1	0.48
oh	1	0.48
wow	1	0.48
place	1	0.48
amaz	1	0.48
…	1	0.48
love	2	0.18
wait	1	0.48
hour	1	0.48
absolut	1	0.48
phone	1	0.48
batteri	1	0.48
last	1	0.48
day	1	0.48

Step 3: Compute TF-IDF

Multiply the TF and IDF values for each term to get the TF-IDF scores.

Document 1

- food: 1/4×0.18=0.045
- delici: 1/4×0.48=0.12
- servic: 1/4×0.48=0.12
- excel: 1/4×0.48=0.12

Document 2

- oh: 1/9×0.48=0.053
- wow: 1/9×0.48=0.053
- place: 1/9×0.48=0.053
- amaz: 1/9×0.48=0.053
- …: 1/9×0.48=0.053
- love: 1/9×0.18=0.02
- wait: 1/9×0.48=0.053

- hour: $1/9 \times 0.48 = 0.053$
- food: $1/9 \times 0.18 = 0.02$

Document 3

- absolut: $1/6 \times 0.48 = 0.08$
- love: $1/6 \times 0.18 = 0.03$
- phone: $1/6 \times 0.48 = 0.08$
- batteri: $1/6 \times 0.48 = 0.08$
- last: $1/6 \times 0.48 = 0.08$
- day: $1/6 \times 0.48 = 0.08$

These TF-IDF scores indicate the importance of each term within its document, taking into account **how common or rare the term is across all documents**.

• df(**t**): Document Frequency (df) is the number of documents containing the term t.

• IDF(**t**): Inverse Document Frequency (IDF) quantifies the importance of the term. If a term appears in many documents, it is *less* significant. Conversely, if it appears in fewer documents, it is *more* informative and significant

- Document 1: "The food is good."
- Document 2: "The food is bad."
- Document 3: "The food is delicious."