TFIDF - Term Frequency - Inverse Document Frequency

D1	My network strength is amazing and i am happy with the service
D2	The customer support is worst here and network is bad
D3	Network strength stays great in city and network works well
D4	Speed of data is good and i am happy with it

Log(4/3) = 0.287682072451781
Log(4/2)=0.693147180559945
Log(4/1) = 1.386294361119891

$$w_{ij} = tf_{i,j} * \log(\frac{N}{df_i})$$

Features	Network	Strength	Amazing	Happy	Service	Customer	Support	Worst	Bad	Great	Well	Speed	Data	Good
DF	3	2	1	2	1	1	1	1	1	1	1	1	1	1
D1	0.287	0.69	1.38	0.69	138	0	0	0	0	0	0	0	0	0
D2	0.28	0	0	0	0	1.38	1.38	1.38	1.38	0	0	0	0	0
D3	0.56	0.69	0	0	0	0	0	0	0	1.38	1.38	0	0	00
D4	0	0	0	1.38	0	0	0	0	0	0	0	1.38	1.38	1.38

TF - IDF = Term Frequency - Inverse Document Frequency

D1	My network strength is amazing and i am happy with the service
D2	The customer support is worst here and network is bad
D3	Network strength stays great in city and network works well
D4	Speed of data is good and i am happy with the service

Log(4/3) = 0.287682072451781
Log(4/2) = 0.693147180559945
Log(4/1)=1.386294361119891

$$w_{ij} = tf_{i,j} * \log(\frac{N}{df_i})$$

Features	Network	Strength	Amazing	Happy	Service	Customer	Support	Worst	Bad	Great	Well	Speed	Data	Good
DF	3	2	1	2	2	1	1	1	1	1	1	1	1	1
D1	0.28	0.69	1.38	0.69	0.69	0	0	0	0	0	0	0	0	0
D2	0.28	0	0	0	0	1.38	1.38	1.38	1.38	0	0	0	0	0
D3	0.56	0.69	0	0	0	0	0	0	0	1.38	1.38	0	0	0
D4	0	0	0	0.69	0.69	0	0	0	0	0	0	1.38	1.38	1.38

45/2=22.5
Log(1/5)=-1.6094379124341

I love python programming

n gram

n=1 unigram → I, love, python, programming

n=2 bigram → I, love, python, programming, I love, love python, python programming

n=3 trigram → I, love, python, programming, python programming, I love python, love python prog- python programming, I love python, love python prog-

ngram - feature extraction method
doc :- I love python programming.

ngram - '

doc :- I love python program 0

n=1 ⇒ one word as a feature

n=1 → unigram → I, love, python, programming.

n=2 → bigram → I, love, python, programming, I love,
love python, python programming.

n=3 → trigram → I, love, python, programming, I love, love python,
python programming, I love python, love python program-

LDA

		Network	Poor	Hate	Bad	Internet	Connection	Strength
		T1	T3	T2	T3	T2	T1	T1

		Network
D1	My network is poor and i hate the network	2 |4 1|4 1 |4
D2	I am getting bad internet connection	1 |3 1|3 1/3
D3	Internet connection is bad	1 |3 1 |3 1/3
D4	Network strength is poor	2 |4 0|3 1|3

Network → T1 = $\frac{2}{4} \times \frac{2}{4} = \frac{4}{16}$

Network → T2 = $\frac{1}{4} \times \frac{2}{4}$

T3 → $\frac{1}{6} \times \frac{2}{4}$

T1
Network - 2/4
connection - 2/4
strength - 1/4

T2
hate - 1/3
internet - 2/3

T3
Poor - 2/4
bad - 2/4