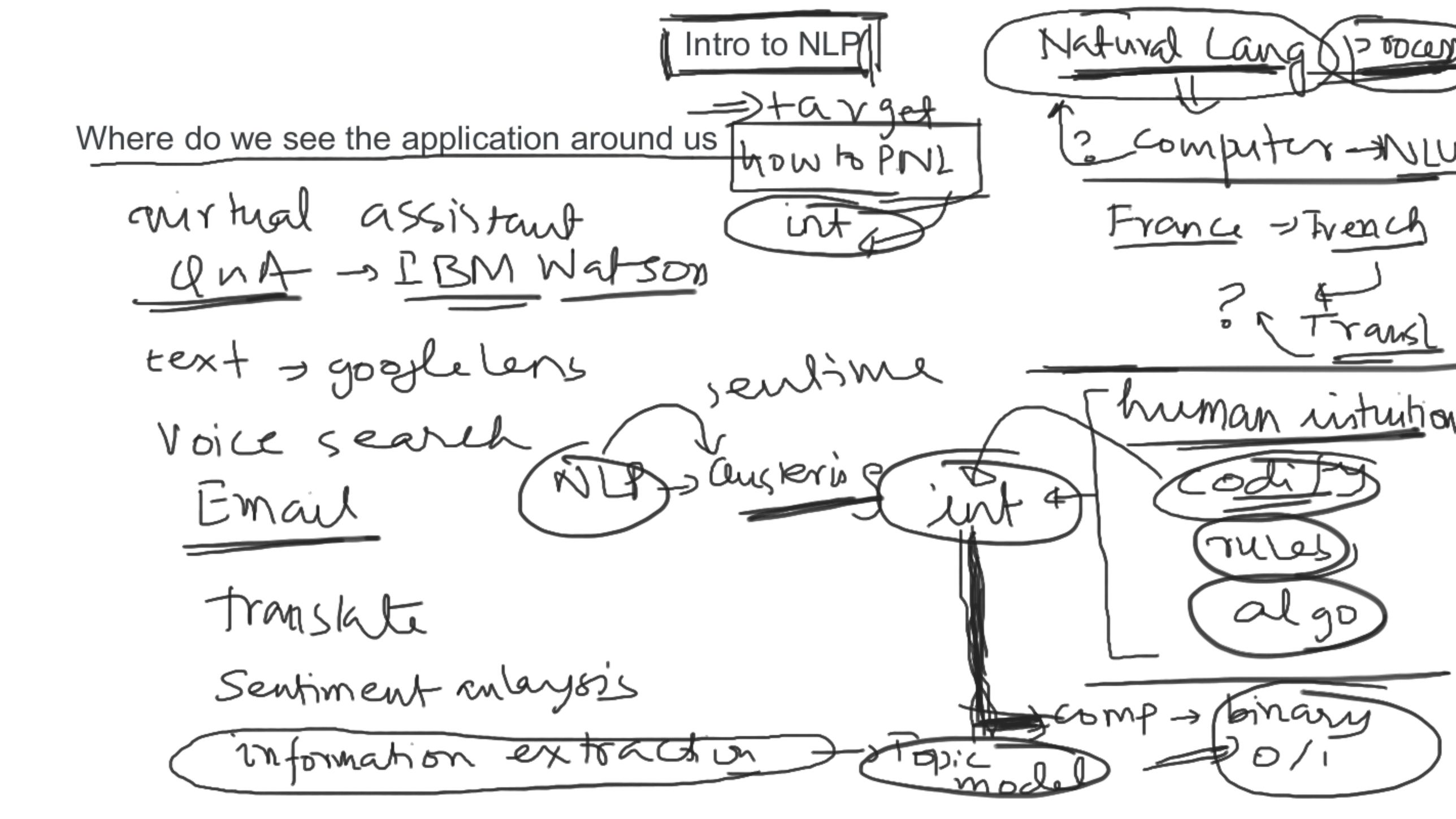

Intro to NLP
target
How to PNL
int
Where do we see the application around us
virtual assistant
QnA → IBM Watson
text → google lens
Voice search
Email
translate
Sentiment analysis
Information extraction
Topic model
NLP → Questions
sentime
int
Natural Lang
computer → NLU
France → French
Transl
human intuition
codify
rules
algo
comp → binary
0/1

DOCUMENT CLUSTERING

- Has applications in knowledge management and information retrieval
- Makes it easy to group similar documents into meaningful groups.

PATTERN IDENTIFICATION

- Process of searching large amounts of text for patterns and recognizing features
- Helps extract features such as phone numbers and email addresses

PRODUCT INSIGHTS

- Helps to extract large amounts of text, such as customer reviews
- Can reveal insights about products

SECURITY MONITORING

- Monitors and extracts information from news articles and reports for national security purposes

Text mining allows you to make better informed decisions, automate information-intensive processes, gather business critical insights, and mitigate operational risks.

Reco content

-ve

I like it.....I can't put it down....but after buying the full set with Rs2110....the 4th book's 3 pages were out and the 5th book which should be the most thick came out to be less thick than the 4th one and the 4th book condition was so bad that I had to put cellotape

The books are fine but definitely not hardcover and what it shows like in the picture. Feels like a scam.

-ve

+ve

The books are great! However, the font is too small and if put to regular reading, the cover starts wearing off - mine did in a month for Order of Phoenix. Also, as the photo shows, one of the pages of the books is missing - Deathly Hallows, page 17. Make sure you check once you've purchased the books! Lol disappointed.

~ 100% X

As everyone knows, this is the popular set of books. I bought this as a gift to my daughter. She is very happy about it. No need to review in detail.

+ve

I love harry Potter series and this book was actually my turning point in reading not that I did not read before that but I was attracted to series and trilogy after reading this ...now coming to the shipping...it was amazing since it came in one day time although I received my Order of Phoenix with a slight cut but it does look noticeable but never mind

These books came in perfect condition.. But the box had a little tear which is not a problem as the box might have been thrown a bit in the delivery so overall it is a good product but it can be made better 🙂

Absolutely loved it ,Good condition book with catchy illustration, Don't read the negative reviews just buy it you will love it 😍

+ve

Packing is terrible. However it's good condition 😔😊

-ve

Till now the service has been excellent....no chance of a single complaint. But not now. I bought this ₹8000+ costing box set of Harry Potter. The outer box was broken....and all the books were cellotaped to stay together. The first two parts of Harry Potter look as if they might have been previously used by someone...there are those marks on the main cover. Secondly, someone has scribbled lines with blue ballpen on the last page of The Chamber of Secrets. It has also got the used marks on the cover. It's completely unexpected. I did not feel like returning it all...as the rest of the books are in good condition. But it does cross my mind that I could have gone for a cheaper set.

It would've been a terrible mistake if I hadn't purchased these books for reading because they were very popular and I thought I would read em. Harry Potter...
You don't know anything If you haven't read it

★☆☆☆☆ **Not hardcover.**

Reviewed in India on 7 December 2019

Verified Purchase

The books are fine but definitely not hardcover and what it shows like in the picture. Feels like a scam.

★★☆☆☆ **Missing pages and cover wearing off**

Reviewed in India on 9 October 2018

Verified Purchase

The books are great! However, the font is too small and if put to regular reading, the cover starts wearing off - mine did in a month for Order of Phoenix. Also, as the photo shows, one of the pages of the books is missing - Deathly Hallows, page 17. Make sure you check once you've purchased the books! Lol disappointed.

★★★★★ **No need of a review on this...**

Reviewed in India on 13 July 2020

Verified Purchase

As everyone knows, this is the popular set of books. I bought this as a gift to my daughter. She is very happy about it. No need to review in detail.

★★★★☆ **Nice books**

Reviewed in India on 21 February 2022

Verified Purchase

These books came in perfect condition.. But the box had a little tear which is not a problem as the box might have been thrown a bit in the delivery so overall it is a good product but it can be made better 🙂

★★★★☆ **Could tail india terrible packing, cover is half open**

Reviewed in India on 11 September 2021

Verified Purchase

Packing is terrible. However it's good condition 😔😉

★★★☆☆ **Good and evil:eternal battle.....**

Reviewed in India on 7 October 2018

Verified Purchase

I like it.....I can't put it down....but after buying the full set with Rs2110....the 4th book's 3 pages were out and the 5th book which should be the most thick came out to be less thick than the 4th one and the 4th book condition was so bad that I had to put cellotape

★★★★☆ **Loved the series**

Reviewed in India on 31 January 2018

Verified Purchase

I love Harry Potter series and this book was actually my turning point in reading not that I did not read before that but I was attracted to series and trilogy after reading this ...now coming to the shipping...it was amazing since it came in one day time although I received my Order of Phoenix with a slight cut but it does look noticeable but never mind

★★★★★ **Loved it**

Reviewed in India on 26 August 2020

Verified Purchase

Absolutely loved it ,Good condition book with catchy illustration, Don't read the negative reviews just buy it you will love it 😍

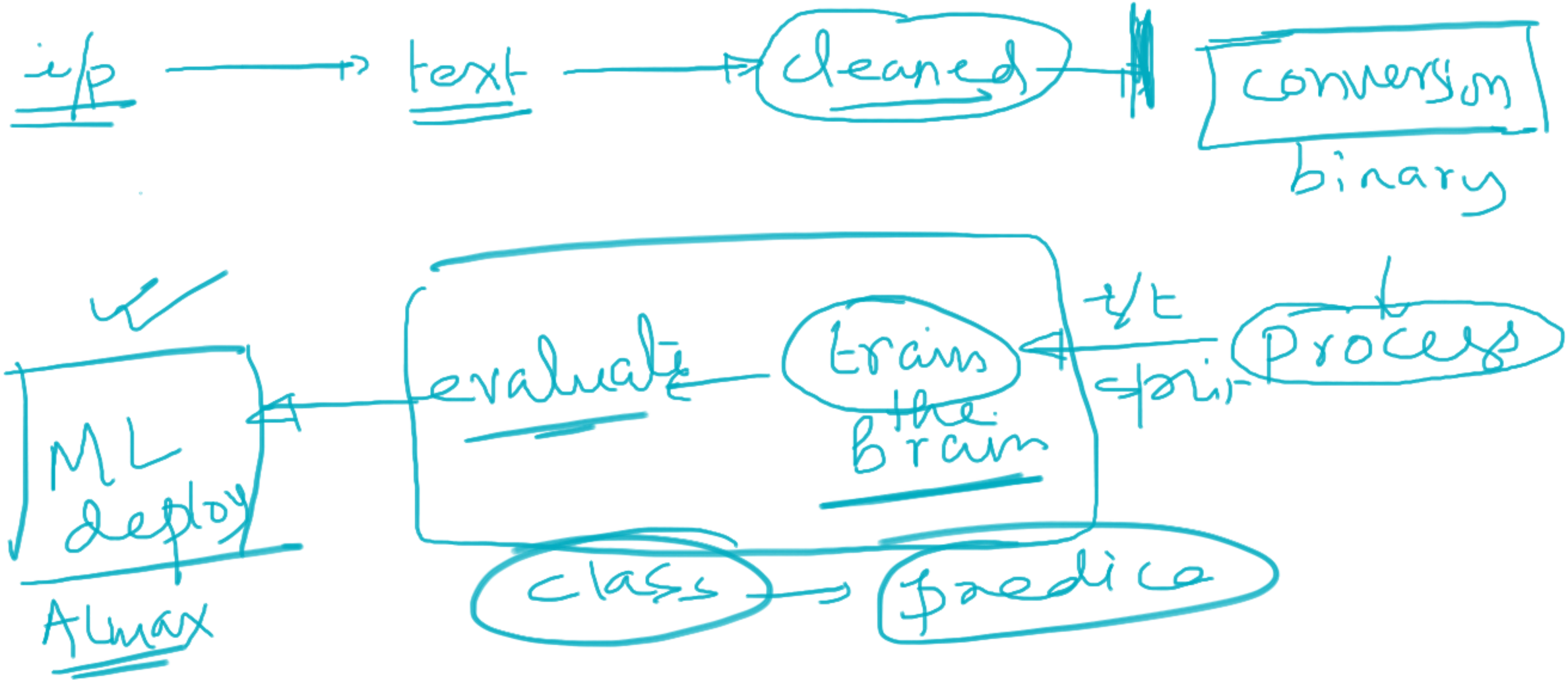
★★★☆☆ **Partly satisfied**

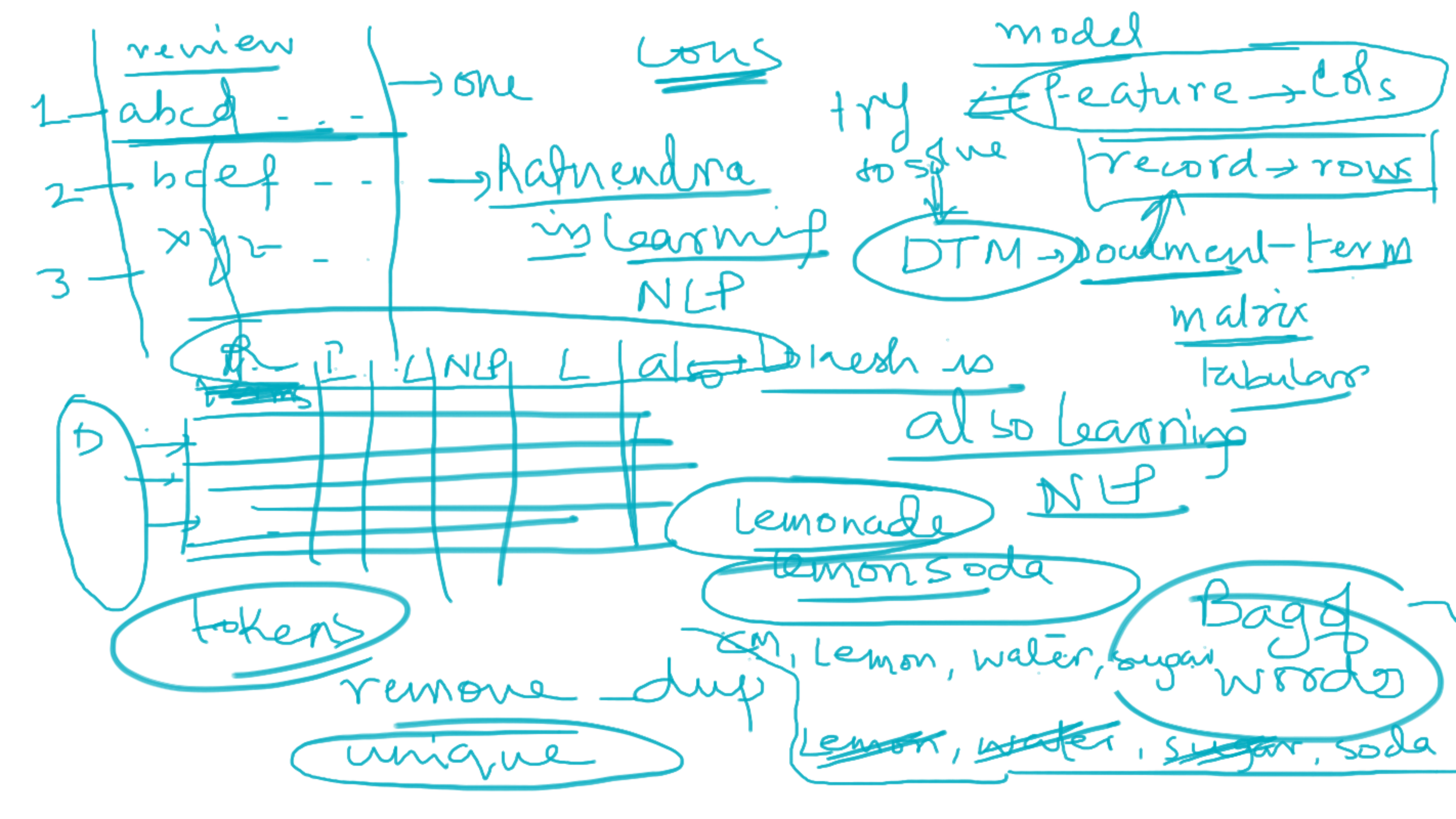
Reviewed in India on 15 July 2021

Verified Purchase

Till now the service has been excellent....no chance of a single complaint. But not now. I bought this ₹8000+ costing box set of Harry Potter. The outer box was broken....and all the books were cellotaped to stay together. The first two parts of Harry Potter look as if they might have been previously used by someone...there are those marks on the main cover. Secondly, someone has scribbled lines with blue ballpen on the last page of Chamber of Secrets. It has also got the used marks on the cover. It's completely unexpected. I did not feel like returning it all....as the rest of the books are in good condition. But it does cross my mind that I could have gone for a cheaper set.

It would've been a terrible mistake if I hadn't purchased these books for reading because they were very popular and I thought I would read em. Harry Potter...
You don't know anything If you haven't read it


i/p
text
cleaned
conversion
binary
w/
ML
deploy
Almax
evaluate
train
the
brain
t/t
split
process
class
predict


review
1 abcd - - -
2 bcef - -
3 xyz - -
one
Lons
Rahnendra
is learning
NLP
model
feature → cols
record → row
try to save
DTM → document-term matrix
tabular
D
I
L
NLP
L
also
Iresh is also learning NLP
tokens
remove dup
unique
Lemonade
Lemon soda
Lemon, water, sugar, soda
Bag of words

Term → Important

garbage in
slow
noisy

Review	Sentiment
The book is not that good	-ve
Worst book I have ever read	-ve
Can't wait for the next version of the book	+ve
I don't like this book	-ve
Like this author	+ve
Really good book	+ve
Bad experience	-ve
Complete waste of money	-ve
Don't wait. Just buy it	-ve +ve
Don't want to buy the next version of the book	-ve -ve

Set of intuitions:

→ punctations (rem)
→ small case
→ stop words (rem)

business

Stop words

can't ↔ cannot

synon + stemming + Lemma

★★★★★ **Hit "buy now". I'm serious.**

Reviewed in India on 24 December 2016

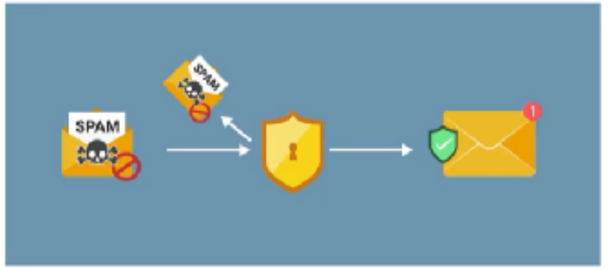
Verified Purchase

It would've been a terrible mistake if I hadn't purchased these books for reading because they were very popular and I thought I would read em. Harry Potter... You don't know anything If you haven't read it

Applications

SPEECH RECOGNITION

SPAM FILTERING



SENTIMENT ANALYSIS

- It determines if a given sentence expresses positive, neutral, or negative sentiments.
- This includes data analysis of the body of text to understand the opinion, modality, and mood.
- It works best on text that has subjective context.

E-COMMERCE PERSONALIZATION

- Text mining is used to suggest products that fit into a user's profile.
- It helps e-commerce retailers learn more about consumers by analyzing information, identifying patterns, and providing insights.
- Retailers can target specific individuals with personalized offers and discounts.
- It increases customer loyalty by identifying purchase patterns and opinions on products.

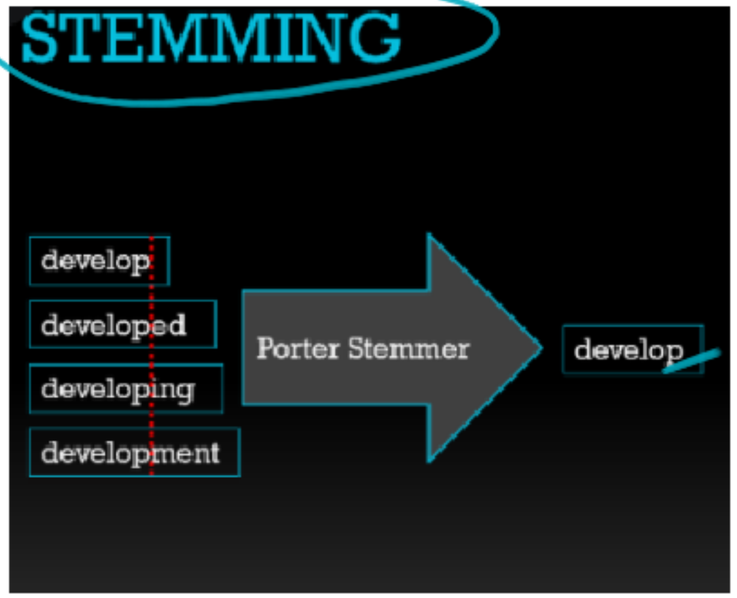
Invent Re-invent

Stemming → Lemma

Grammer

Form	Suffix	Stem
helps	-es	help
helping	-ing	help
helped	-ed	help
help	-	help
helper	-er	help

Removal of affixes



What about the word "Drive"

help

f_1 f_2 f_3 f_n f_5

f_1

set of rules

suffix → root word

Stemming algorithms

- Porter Stemmer: Most common and gentle stemmer.
- Fast, imprecise.

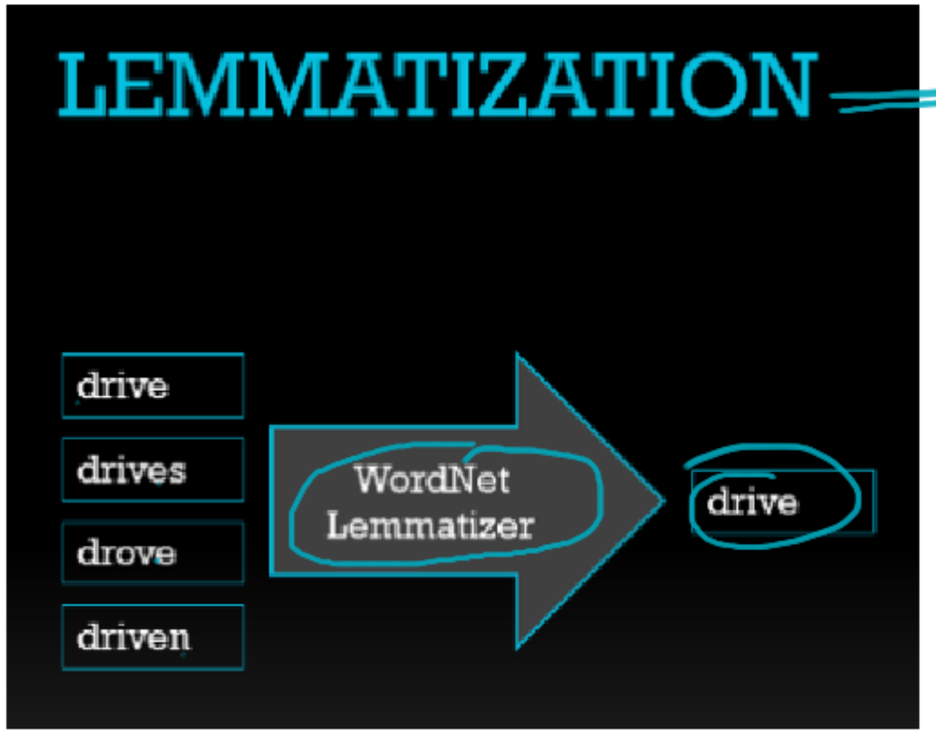
Stemming algorithms

- Snowball Stemmer (Porter2 / English stemmer): preciseness targeted over large datasets.
- Supports custom rules for any language.

Stemming algorithms

- Lancaster Stemmer (Paice-Husk): AGGRESSIVE. ITERATIVE.
- Simple, but heavy stemming can cause over-stemming.
- Leads to non-linguistic, with no meaning.

Lemmatization



complex → root

dictionary

lot of time → processing high

Stem → stemmed

Lemma → root (meaning)

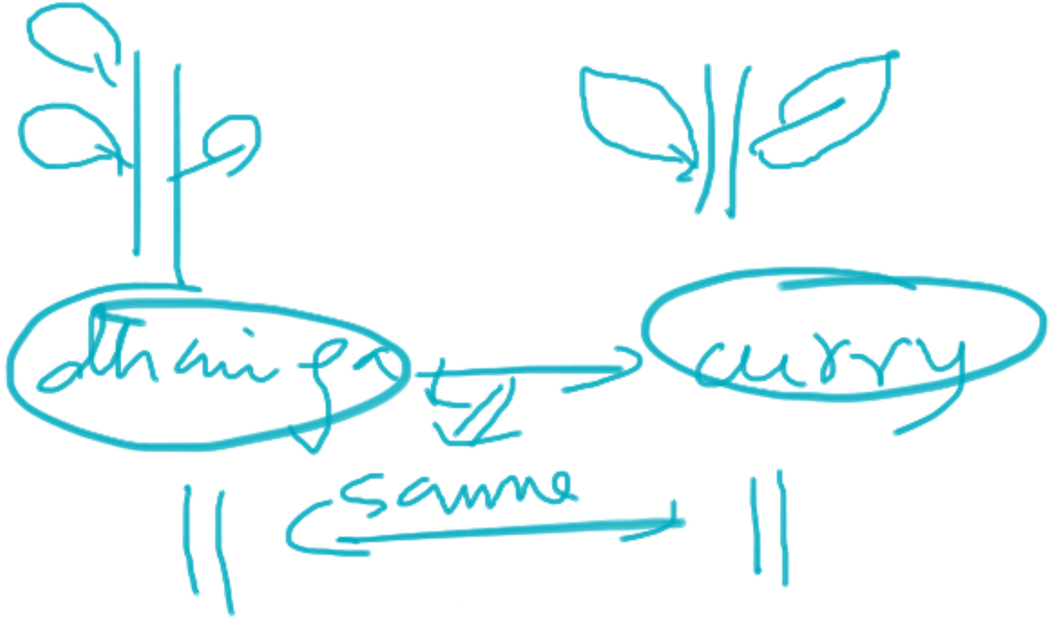
Overstemming

Understemming

"universal", "university", and "universe"

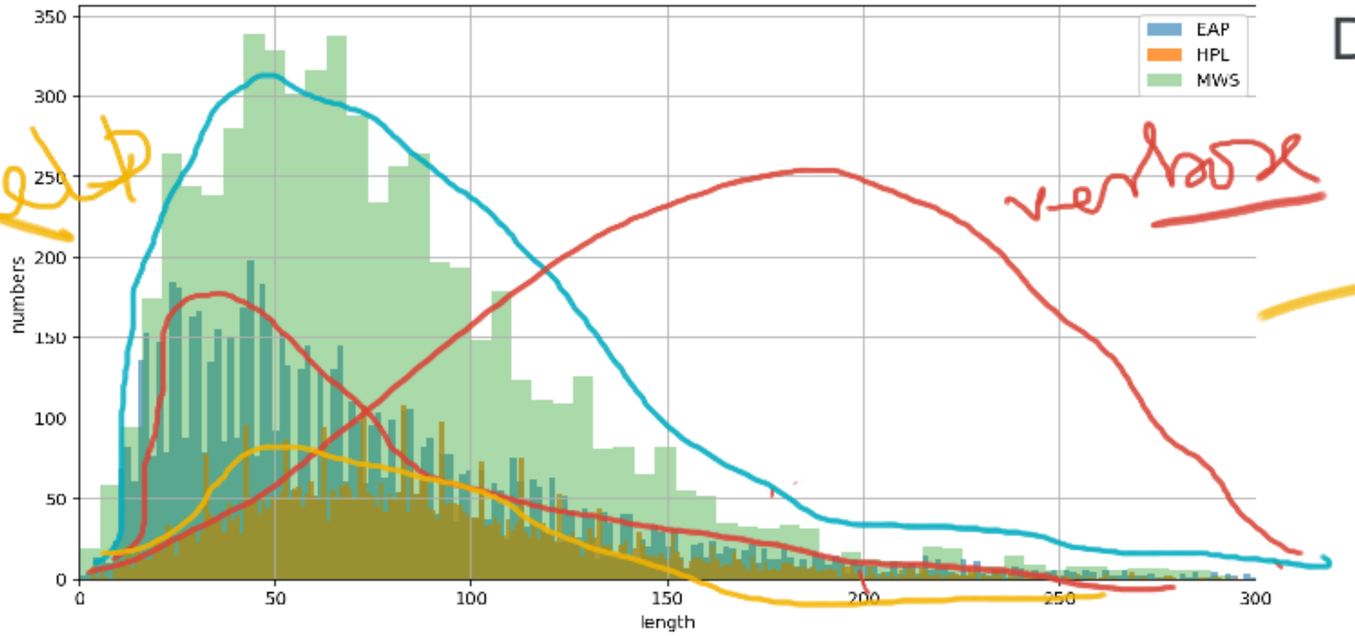
to "univers"

"alumnus" → "alumnu"
"alumni" → "alumni",
"alumna"/"alumnae" → "alumna"



Original	Stemmed	Lemmatized
visibilities	visibl	visibility
adhere	adher	adhere
adhesion	adhes	adhesion
appendicitis	append	appendicitis
oxen	oxen	ox
indices	indic	index
swum	swum	swim

Advantages



Disadvantages

DTM = ? →

Metro / City / Town student sl(e) → els

get_dummies

book

help

Nos. → Lokesh learn NLP

help heal help half Tauf learn

NLP ch Co

M	T	C
1	0	0
0	1	0
0	1	0
1	0	0
0	0	1
0	0	1

M
T
T
M
T
C
C
C
T

Ch	Co	Lok	Learn	NLP	Tauf
0	0	1	1	1	0
1	1	0	1	1	1

count | words | terms

DTM

relevant

book	learn	ch	help
20	1		
25	2	1	
30	3	4	

⟹ countvectorizer

book = how many doc contain book

TF-IDF

Term n

inv doc n

Term n x 1 / doc n

tf

d1: "sky blue

d2: "sun bright today"

d3: "sun sky bright"

d4: "can see shining sun bright sun"

$f_{t,d}$ → n table

	blue	bright	can	see	shining	sky	sun	today
1	1	0	0	0	0	1	0	0
2	0	1	0	0	0	0	1	1
3	0	1	0	0	0	1	1	0
4	0	1	1	1	1	0	2	0

d1: "sky blue

d2: "sun bright today"

d3: "sun sky bright"

d4: "can see shining sun bright sun"

copy

$f_{t,d}$ TF

	blue	bright	can	see	shining	sky	sun	today
1	1	0	0	0	0	1	0	0
2	0	1	0	0	0	0	1	1
3	0	1	0	0	0	1	1	0
4	0	1	1	1	1	0	2	0

IDF

$f_{t,d}$

	blue	bright	can	see	shining	sky	sun	today
1	1	0	0	0	0	1	0	0
2	0	1	0	0	0	0	1	1
3	0	1	0	0	0	1	1	0
4	0	1	1	1	1	0	2	0
n_t	**1**	**3**	**1**	**1**	**1**	**2**	**3**	**1**

step-3 (i):

count one hot

$\log \frac{N}{n_t}$ idf = $\log \frac{n_1}{N}$ → # of rows

$f_{t,d}$

	blue	bright	can	see	shining	sky	sun	today
1	1	0	0	0	0	1	0	0
2	0	1	0	0	0	0	1	1
3	0	1	0	0	0	1	1	0
4	0	1	1	1	1	0	2	0

$\mathrm{tf}(t,d) = \frac{f_{t,d}}{\sum_{t'} f_{t',d}}$

	blue	bright	can	see	shining	sky	sun	today
1	1/2	0	0	0	0	1/2	0	0
2	0	1/3	0	0	0	0	1/3	1/3
3	0	1/3	0	0	0	1/3	1/3	0
4	0	1/6	1/6	1/6	1/6	0	1/3	0

fract → step 3

$\mathrm{idf}(t,D) = \log_{10} \frac{N}{n_t}$

$f_{t,d}$

	blue	bright	can	see	shining	sky	sun	today
1	1	0	0	0	0	1	0	0
2	0	1	0	0	0	0	1	1
3	0	1	0	0	0	1	1	0
4	0	1	1	1	1	0	2	0
n_t	1	3	1	1	1	2	3	1

$N = 4$

	blue	bright	can	see	shining	sky	sun	today
	0.602	0.125	0.602	0.602	0.602	0.301	0.125	0.602

$\log_{10} \frac{4}{1} = 0.602$ $\log_{10} \frac{4}{3} = 0.125$

$N = 10^2$ power

$\log \frac{N}{n_t}$

$\mathrm{tf}(t,d)$

	blue	bright	can	see	shining	sky	sun	today
1	1/2	0	0	0	0	1/2	0	0
2	0	1/3	0	0	0	0	1/3	1/3
3	0	1/3	0	0	0	1/3	1/3	0
4	0	1/6	1/6	1/6	1/6	0	1/3	0

2

X

$\mathrm{idf}(t,D)$

	blue	bright	can	see	shining	sky	sun	today
	0.602	0.125	0.602	0.602	0.602	0.301	0.125	0.602

- TF-IDF: Multiply TF and IDF scores, use to rank importance of words within documents
  - Most important word for each document is highlighted

TF-IDF score computation. [Image Source]

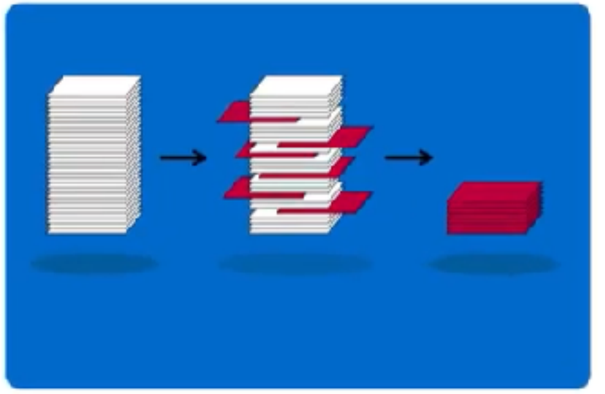
$\mathrm{tfidf}(t,d,D) = \mathrm{tf}(t,d) \cdot \mathrm{idf}(t,D)$

	blue	bright	can	see	shining	sky	sun	today
1	**0.301**	0	0	0	0	0.151	0	0
2	0	0.0417	0	0	0	0	0.0417	**0.201**
3	0	0.0417	0	0	0	**0.100**	0.0417	0
4	0	0.0209	**0.100**	**0.100**	**0.100**	0	0.0417	0

list of doc

imp

It uses software that can identify concepts, patterns, topics, keywords, and so on in the data.

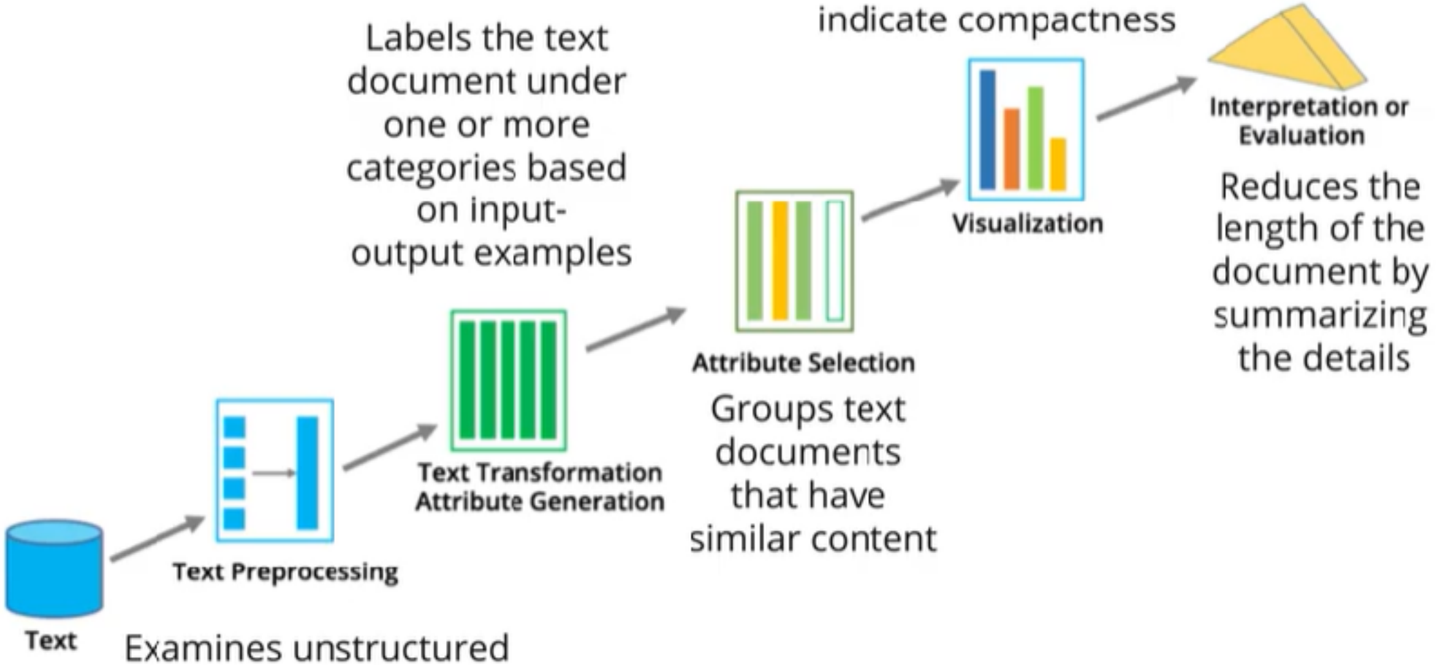


It uses computational techniques to extract high-quality information from unstructured text.

topic model

Flow of Text Mining

	id	text	author
0	id26305	This process, however, afforded me no means of...	EAP
1	id17569	It never once occurred to me that the fumbling...	HPL
2	id11008	In his left hand was a gold snuff box, from wh...	EAP
3	id27763	How lovely is spring As we looked from Windsor...	MWS
4	id12958	Finding nothing else, not even gold, the Super...	HPL
5	id22965	A youth passed in solitude, my best years spen...	MWS
6	id09674	The astronomer, perhaps, at this point, took r...	EAP
7	id13515	The surcingle hung in ribands from my body.	EAP
8	id19322	I knew that you could not say to yourself 'ste...	EAP
9	id00912	I confess that neither the structure of langua...	MWS



NLP Process Workflow

TOKENIZATION

Splits text into pieces (tokens), remove punctuation.

STOPWORD REMOVAL

Removes commonly used words (Such as 'the') which are not relevant to analysis.

STEMMING AND LEMMATIZATION

Reduces words to base form to be analyzed as a single item.

P.O.S TAGGING

Tags words to be part of speech (Such as verb, noun) based on definition and context.

INFORMATION RETRIEVAL

Extracts relevant information from source