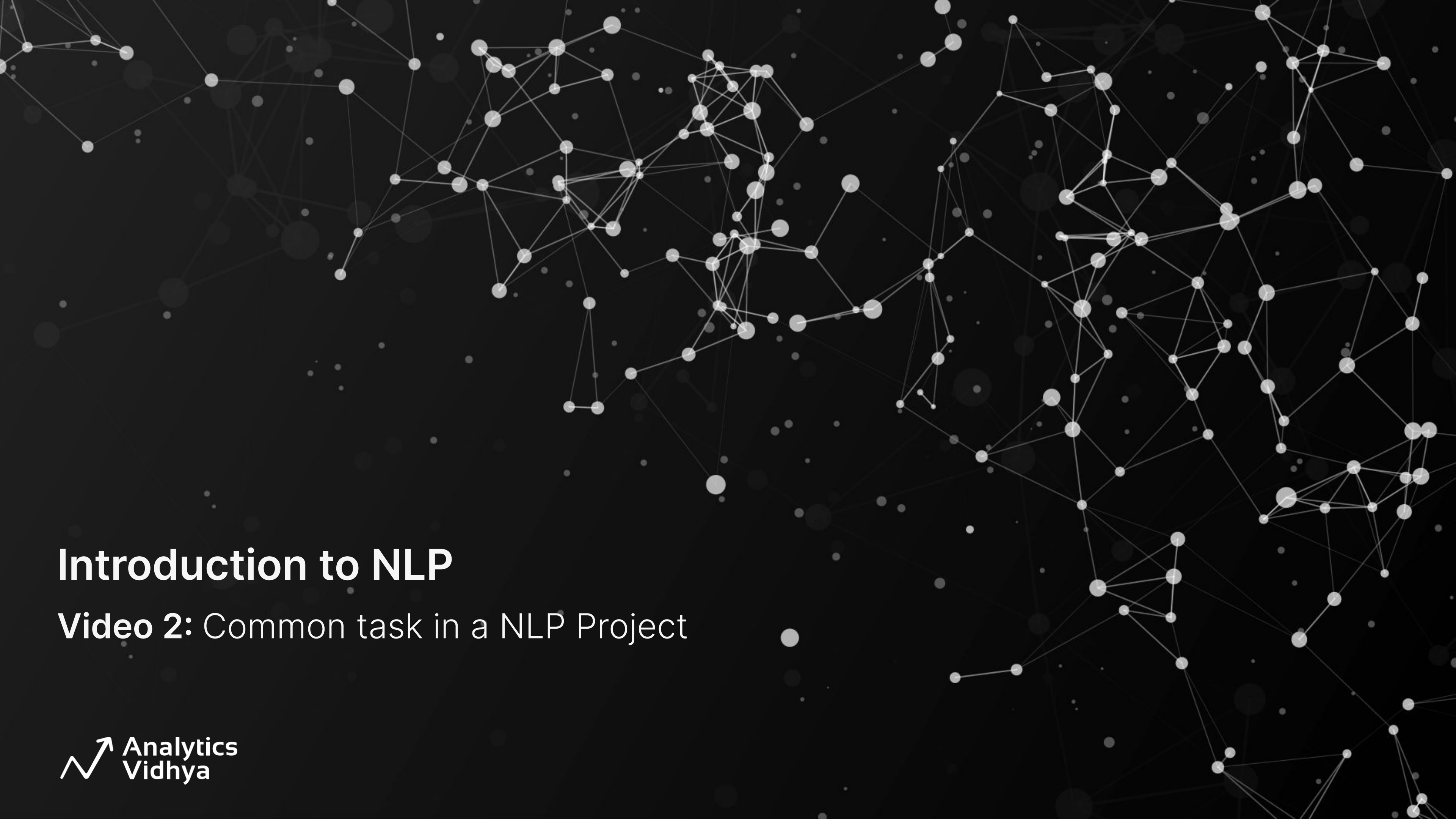
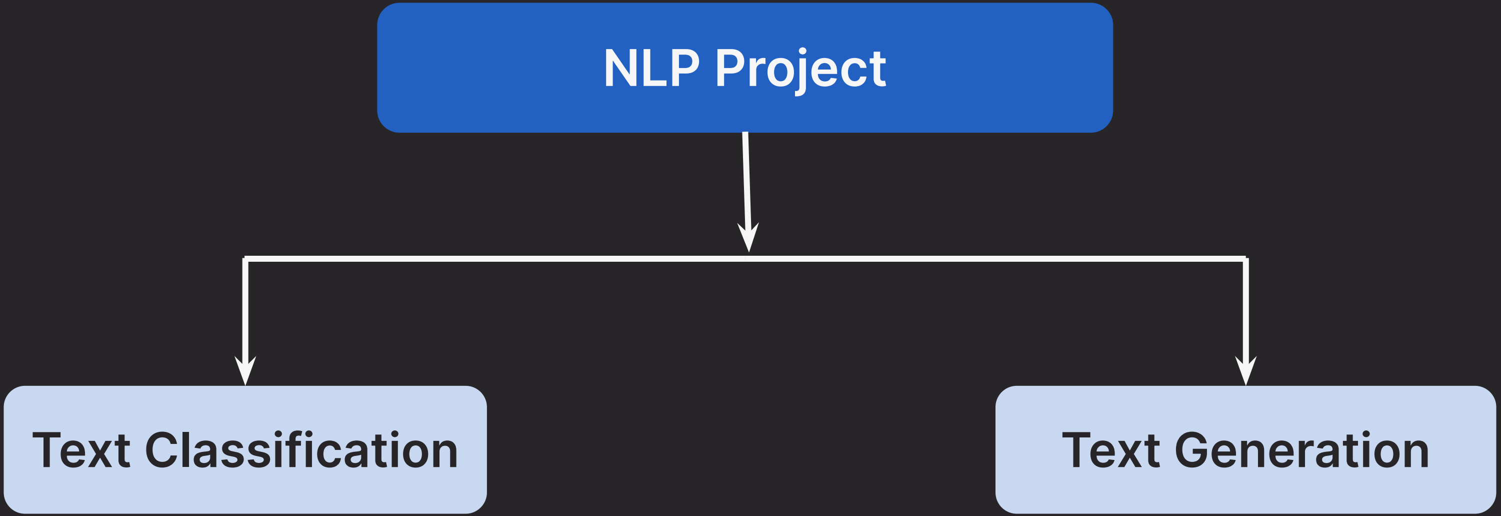


Introduction to NLP

Video 2: Common task in a NLP Project

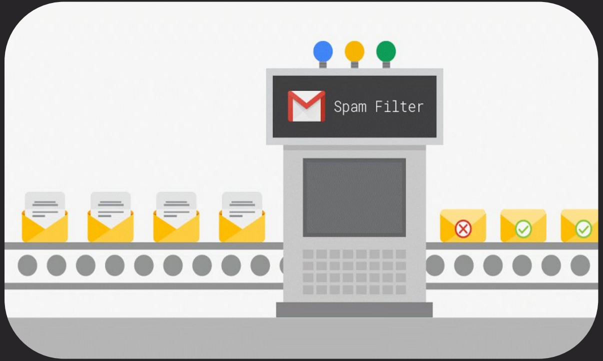
Analytics
Vidhya

Natural Language Processing



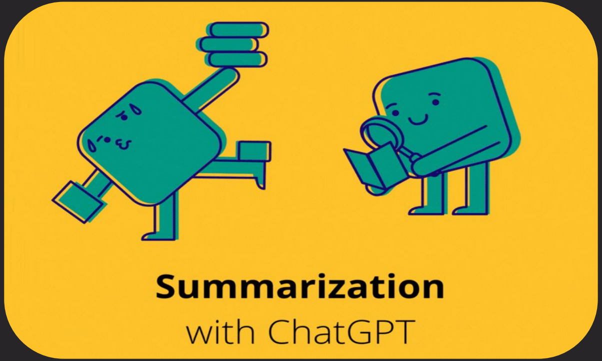
NLP Projects

1. Text Classification



Segregate texts based on the content

2. Text Generation



Generate text based on certain inputs

Techniques in NLP

Common Preprocessing Techniques

Advanced Techniques

Techniques in NLP

Common Preprocessing Techniques

Advanced Techniques

Common Preprocessing Techniques

☑ Count of total occurance

Document	pets	animal	cats	dogs	fish
Cats and dogs are pets	1	0	1	1	0
Wild animals are not pets	1	1	0	0	0
Some people keep fish as pets	1	0	0	0	1

Common Preprocessing Techniques

☑ Count of total occurance

☑ Break text into individual words

Cats

and

dogs

are

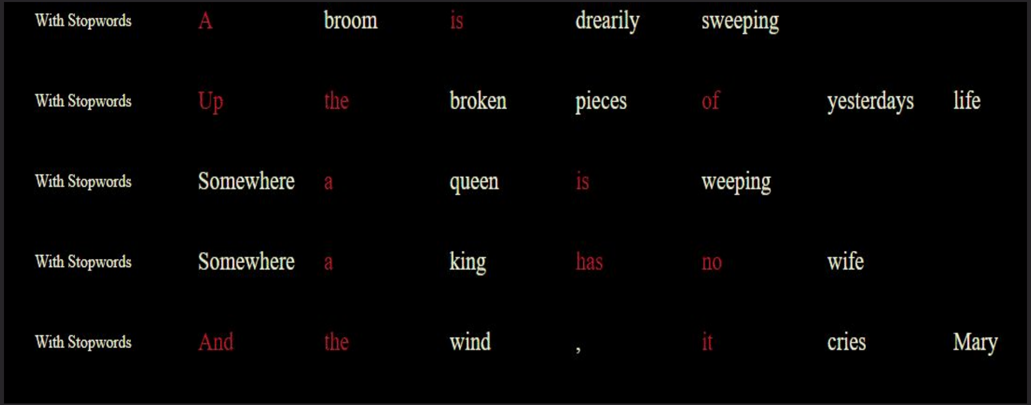
pets

Common Preprocessing Techniques

☑ Count of total occurance

☑ Break text into individual words

☑ Removal of stop words like “a”, “the”, “is” etc.



With Stopwords	A	broom	is	drearily	sweeping		
With Stopwords	Up	the	broken	pieces	of	yesterdays	life
With Stopwords	Somewhere	a	queen	is	weeping		
With Stopwords	Somewhere	a	king	has	no	wife	
With Stopwords	And	the	wind	,	it	cries	Mary

Common Preprocessing Techniques

- ☑ Count of total occurance
- ☑ Break text into individual words
- ☑ Removal of stop words like “a”, “the”, “is” etc.

☑ Convert all text to lowercase

RAW	Lowercase
Cat CAT cAt	cat

Common Preprocessing Techniques

- ☑ Count of total occurance
- ☑ Break text into individual words
- ☑ Removal of stop words like “a”, “the”, “is” etc.
- ☑ Convert all text to lowercase
- ☑ Reduce words to their root form called lemma

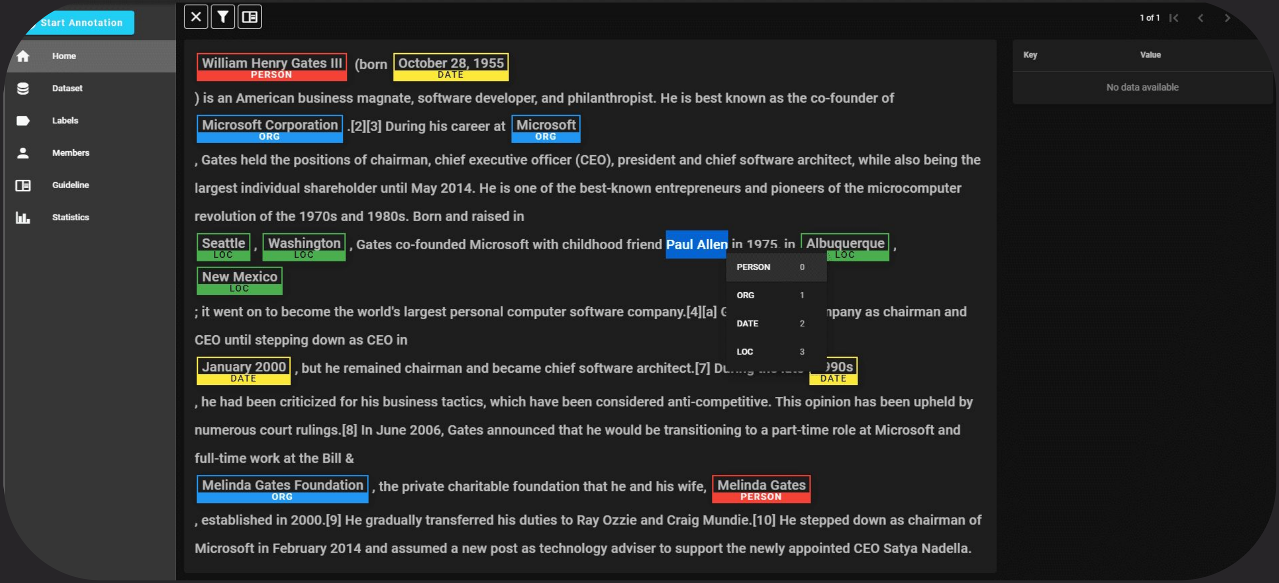
Original word	Lemmatized word
trouble troubling Troubled troubles	trouble

Techniques in NLP

Common Preprocessing Techniques

Advanced Techniques

1. Name Entity Recognition



William Henry Gates III PERSON (born October 28, 1955 DATE

) is an American business magnate, software developer, and philanthropist. He is best known as the co-founder of

Microsoft Corporation ORG .[2][3] During his career at Microsoft ORG

, Gates held the positions of chairman, chief executive officer (CEO), president and chief software architect, while also being the largest individual shareholder until May 2014. He is one of the best-known entrepreneurs and pioneers of the microcomputer revolution of the 1970s and 1980s. Born and raised in

Seattle LOC , Washington LOC , Gates co-founded Microsoft with childhood friend Paul Allen in 1975 in Albuquerque LOC ,

New Mexico LOC

; it went on to become the world's largest personal computer software company.[4][a] G npany as chairman and CEO until stepping down as CEO in

January 2000 DATE , but he remained chairman and became chief software architect.[7] Du 90s DATE

, he had been criticized for his business tactics, which have been considered anti-competitive. This opinion has been upheld by numerous court rulings.[8] In June 2006, Gates announced that he would be transitioning to a part-time role at Microsoft and full-time work at the Bill &

Melinda Gates Foundation ORG , the private charitable foundation that he and his wife, Melinda Gates PERSON

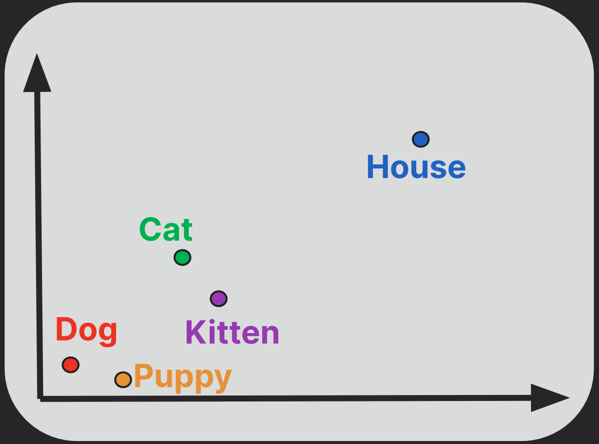
, established in 2000.[9] He gradually transferred his duties to Ray Ozzie and Craig Mundie.[10] He stepped down as chairman of Microsoft in February 2014 and assumed a new post as technology adviser to support the newly appointed CEO Satya Nadella.

2. Word Embeddings

Cat	0.6	0.9	0.1	0.4	-0.7	-0.3	-0.2
Kitten	0.5	0.8	-0.1	0.2	-0.6	-0.5	-0.1
Dog	0.7	-0.1	0.4	0.3	-0.4	-0.1	-0.3
Puppy	-0.8	-0.4	-0.5	0.1	-0.9	0.3	0.8
House	-0.8	-0.4	-0.5	0.1	-0.9	0.3	0.8

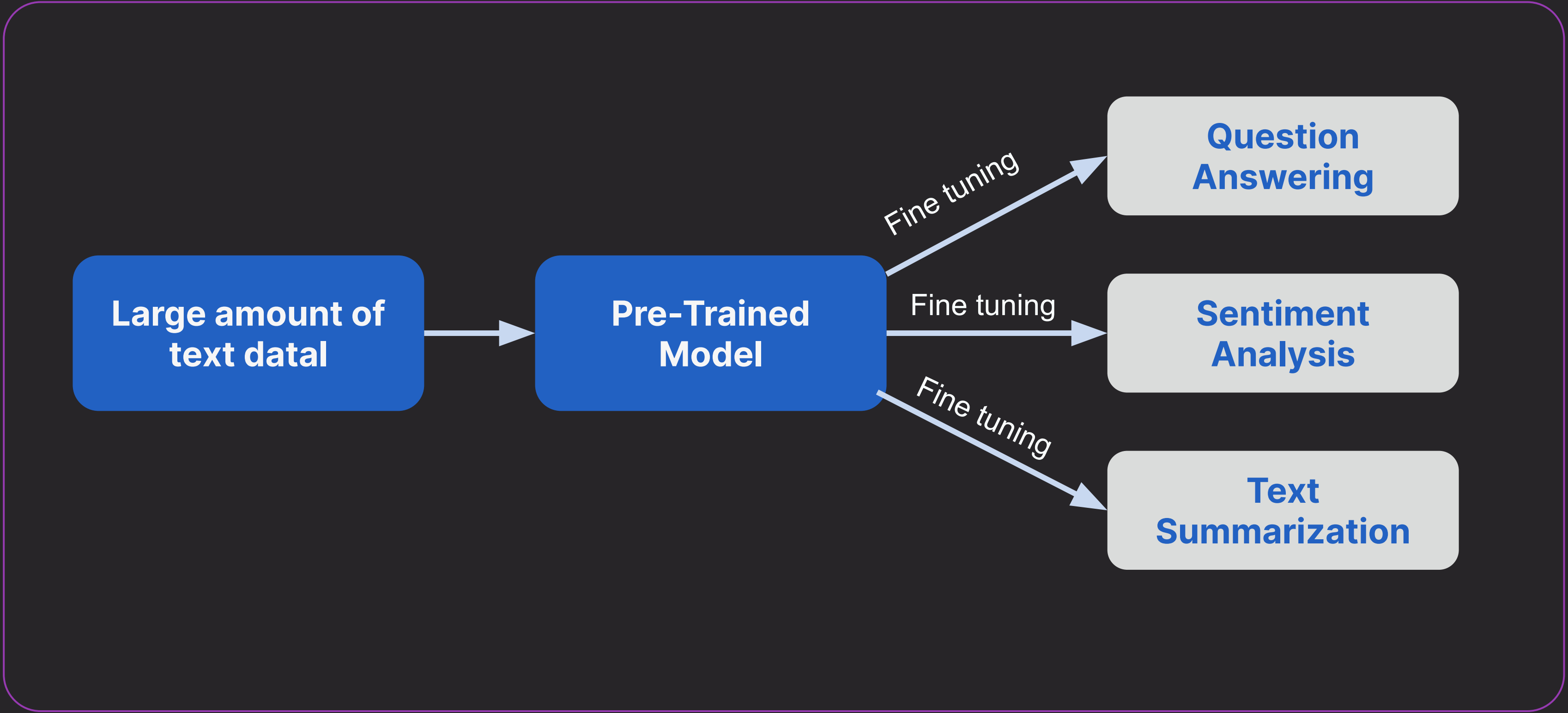
Word

Word Embedding



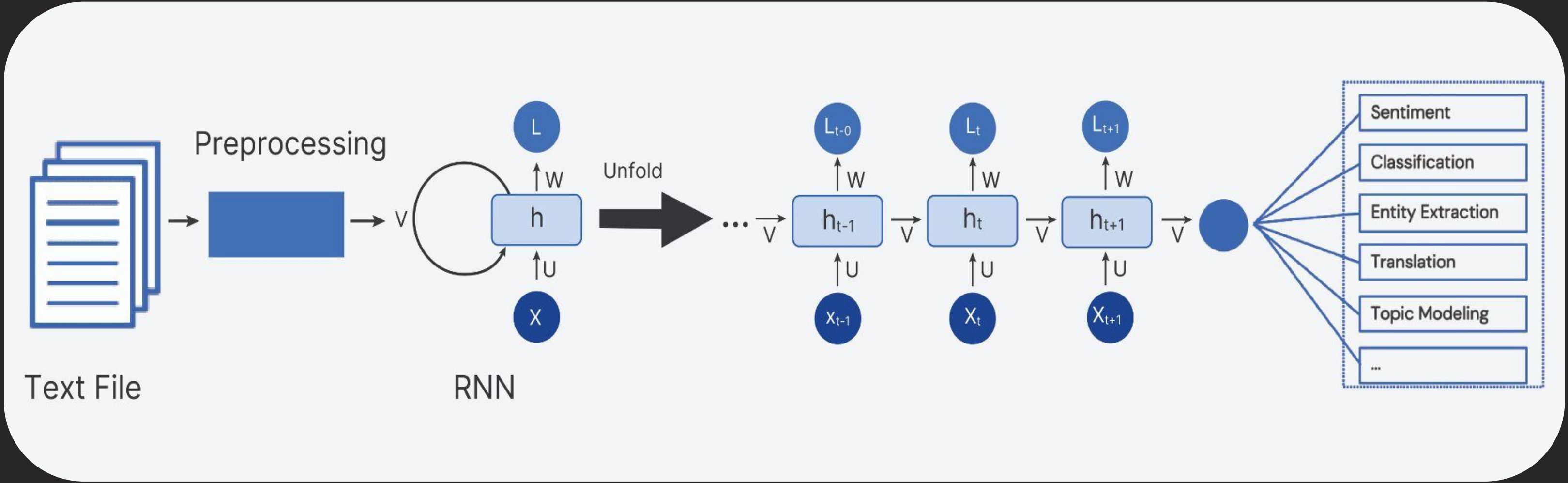
Visualization of word embedding in 2D

3. Fine Tuning Pre-trained models



Analytics Vidhya

4. Building Deep learning models



Techniques in NLP

Common Preprocessing Techniques

Advanced Techniques