

One Hot Encoding

Steps Involved

- Text Cleaning



Steps Involved

- Text Cleaning
- Tokenization



Steps Involved

- Text Cleaning
- Tokenization
- Vocabulary Preparation



Steps Involved

- Text Cleaning
- Tokenization
- Vocabulary Preparation
- One Hot Encoded Vectors



Text Preprocessing

Raw Text
The ball is pink.
Pass the juice.
The book is interesting.

Text Cleaning

Raw Text	Cleaned Text
The ball is pink.	the ball is pink
Pass the juice.	pass the juice
The book is interesting.	the book is interesting

Tokenization

Cleaned Text	Tokens
the ball is pink	the, ball, is, pink
pass the juice	pass, the, juice
the book is interesting	the, book, is, interesting

Vocabulary Preparation

Tokens
the, ball, is, pink
pass, the, juice
the, book, is, interesting



ball, book,
interesting, is, juice,
pass, pink, the

Vocabulary

One Hot Encoded Vectors

ball



One Hot Encoded Vectors

ball $\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$



One Hot Encoded Vectors

ball	[1	0	0	0	0	0	0	0]
book	[0	1	0	0	0	0	0	0]



One Hot Encoded Vectors

ball	[1	0	0	0	0	0	0]
book	[0	1	0	0	0	0	0]
interesting	[0	0	1	0	0	0	0]
is	[0	0	0	1	0	0	0]
juice	[0	0	0	0	1	0	0]
pass	[0	0	0	0	0	1	0]
pink	[0	0	0	0	0	0	1]
the	[0	0	0	0	0	0	1]

 Thank You 
Analytics Vidhya