One Hot Encoding S



Text Cleaning





- Text Cleaning
- Tokenization





- Text Cleaning
- Tokenization
- Vocabulary Preparation





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





Text Preprocessing

Raw Text

The ball is pink.

Pass the juice.

The book is interesting.

Analytics Vidhya



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



ball

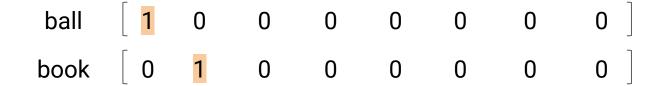




ball 1 0 0 0 0 0 0 0

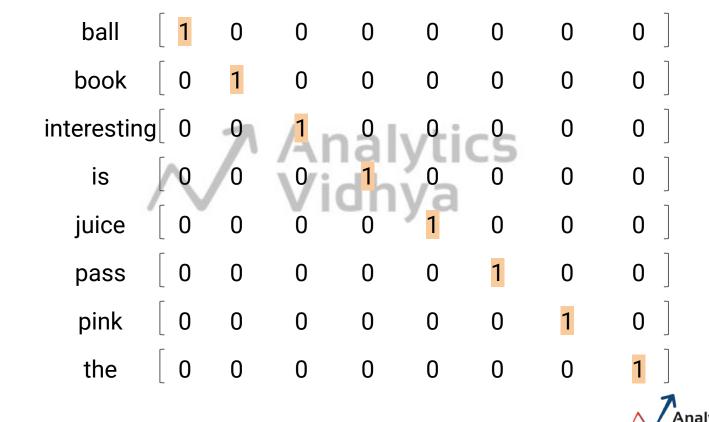












Thank You tics Vidhya

