

# Text Feature Engineering Techniques

# Different Text Feature Engineering Techniques

- Special characters and numbers



# Text Feature Engineering: Special Chars and Num

- Special characters and numbers
  - > Pichai, Mr. Sundar



# Text Feature Engineering: Special Chars and Num

- Special characters and numbers

> I love @AnalyticsVidhya. They are a true #dataanalytics and #datascience community!



# Text Feature Engineering: Special Chars and Num

- Special characters and numbers

> We offer the best pastries in the city! All you have to do is pick up the phone and call 9897xxxxxx!



# Text Feature Engineering: Special Chars and Num

- Special characters and numbers

And so on!



# Different Text Feature Engineering Techniques

- Special characters and numbers
- Word count



# Text Feature Engineering: Word Count

Sentence	WordCount
I love this beautiful city	5
This is a pathetic restaurant that serves you terrible and indigestible dishes that will suck out the flavour from all your future meals	23



# Different Text Feature Engineering Techniques

- Special characters and numbers
- Word count
- Average word length



# Text Feature Engineering: Average Word length

Sentence	Average word len
Exasperating farrago of distortions, misrepresentations and outright lies being broadcasted	8.2
First 5k today! Felt great! OSM!	4.5

# Different Text Feature Engineering Techniques

- Special characters and numbers
- Word count
- Average word length
- Stopwords



# Text Feature Engineering: Stopwords

Text	Stopword count
We need your support to turn this dream into a reality	6

# Different Text Feature Engineering Techniques

- Special characters and numbers
- Word count
- Average word length
- Stopwords
- POS tags and NER



# Text Feature Engineering: POS and NER

Text	NER	PERSON	NOUN	VERB
The 2007 T20 World Cup redefined the contours of Indian cricket helping the young blood, under the leadership of MS Dhoni, rise to the top.	4	1	5	3

# Different Text Feature Engineering Techniques

- Special characters and numbers
- Word count
- Average word length
- Stopwords
- POS tags and NER
- And more!





Thank You!