

## DAFTAR ISTILAH

Istilah	Deskripsi	Halaman pertama kali digunakan
<i>Supervised learning</i>	: Algoritma yang memunculkan suatu fungsi yang akan memetakan <i>input</i> ke <i>output</i> yang diinginkan	3
<i>Bag-of-words</i>	: Model yang diterapkan pada kategorisasi teks, pengelompokan dokumen, atau pengambilan informasi, beroperasi pada tingkat dokumen	11
Token	: Kemunculan angka, huruf, atau kata yang dipisahkan dengan spasi	14
<i>Database</i>	: Transformasi data yang tidak terstruktur seperti HTML atau XML yang ada di web menjadi data yang terstruktur	15
<i>Incomplete</i>	: Kurangnya atribut pada suatu data	16
<i>Noisy</i>	: Adanya <i>error</i> dan <i>outliers</i> pada data	16
<i>Inconsistent</i>	: Adanya inkonsistensi pada data seperti penamaan, singkatan, ataupun kode	16
<i>Lemma</i>	: Perubahan kata berdasarkan bentuk dasar atau kamus, bisa dilakukan dengan cara menghapus akhiran atau awalan, seperti bentuk tunggal pada kata benda, bentuk infinitif pada kata kerja.	20
<i>Term</i>	: Frasa atau kata menjadi predikat atau subjek dari sebuah proposisi	21