GLOSSARY

NLP with Python for Machine Learning Essential Training

With Derek Jedamski

Use these terms and definitions below to understand concepts taught in the course.

Transcript Search: note that you can search for terms directly within the course. To search video text, switch to the *Transcripts* tab, then press Cmd/Ctrl + F on your keyboard to run a search within the active transcript.

Term	Definition
feature engineering	The process of creating new features and/or transforming existing features to get the most out of your data
lemmatizing	The process of grouping together the inflected forms of a word so they can be analyzed as a single term
root cause	The earliest underlying factor contributing to a problem
NLP	Acronym for “Natural Language Processing” that refers to the field concerned with the ability of a computer to understand, analyze, manipulate, and potentially generate human language
NLTK	Acronym for “natural language toolkit” and is the most utilized package for handling natural language processing tasks in Python; open-source tools created for the purpose of making NLP process in Python easier
regular expressions	A text string used for describe a certain search pattern; also known as a “regex”
stemming	The process of reducing inflection or derived words to their word stem or root; crudely chopping off the ends of a word to leave only the base
stop words	Commonly used words like “but,” “if,” “the,” etc.
tokenizing	The process of splitting some string or sentence into a list of words
vectorizing	The process of encoding text as integers to create feature vectors