GLOSSARY

Data Science Foundations: Fundamentals

With Barton Poulson

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
API	Application programming interface. A tool that allows data scientists to access structured data from a database or other repository
Bayes probability	A data analysis technique that uses prior probabilities and test information to create posterior (after the fact) probabilities
big data	An amount of data that is so large that it cannot be processed by a single computer. Analysis of big data can reveal important insights and trends not possible with smaller sample sets.
naive Bayes classifiers	A method of classifying machine learning algorithms, using the data given. "Naive" Bayes classifiers ignore relationships between predictors.
open data	Data that is freely available for noncommercial and/or commercial use. Open data may be inaccurate, biased, or non-anonymized.
scraping	A technique for accessing data from a website or repository when an API is not available, using tools like import.io and scripting