What is Machine Learning?

Typing "what is machine learning?" into a Google search opens up a pandora's box of forums, academic research, and false information – and the purpose of this article is to simplify the definition and understanding of machine learning thanks to the direct help from our panel of machine learning researchers.

At Emerj, the AI Research and Advisory Company, many of our enterprise clients feel as though they should be investing in machine learning projects, but they don't have a strong grasp of what it is. We often direct them to this resource to get them started with the fundamentals of machine learning in business.

What is TfidfVectorizer?

Without going into the math, TF-IDF are word frequency scores that try to highlight words that are more interesting, e.g. frequent in a document but not across documents. The TfidfVectorizer will tokenize documents, learn the vocabulary and inverse document frequency weightings, and allow you to encode new documents

What is PorterStemmer?

Over-stemming occurs when two words are stemmed from the same root that are of different stems. Over-stemming can also be regarded as false-positives. Under-stemming occurs when two words stem from the same root that are not of different stems.