NLP Classification

Text Classification

It’s an instance of classification/categorization where the input data point(s) is text and the goal is to categorize the piece of text into one or more buckets (called a class) from a set of pre-defined buckets (classes)

Applications

*Content classification and organization*

* task of classifying/tagging large amounts of textual data
* tagging product descriptions in an e-commerce website
* routing customer service requests in a company to the appropriate support team
* organizing emails into personal, social, and promotions in Gmail

*Customer support*

* Customers often use social media to express their opinions about and experiences of products or services. Text classification is often used to identify the tweets that brands must respond to (i.e., those that are actionable) and those that don’t require a response



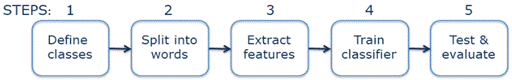
*E-commerce*

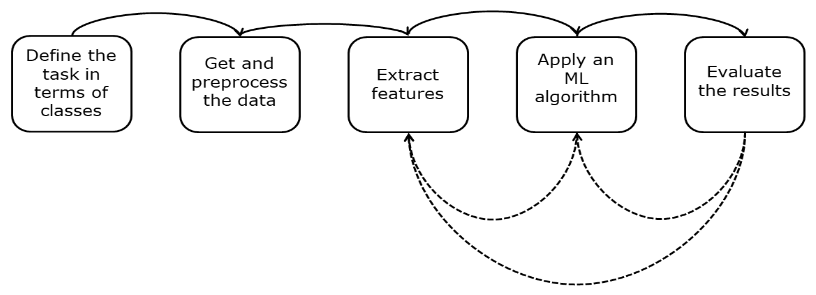
Customers leave reviews for a range of products on e-commerce websites like Amazon, eBay, etc. An example use of text classification in this kind of scenario is to understand and analyze customers’ perception of a product or service based on their comments. This is commonly known as “sentiment analysis.”

Spam detection

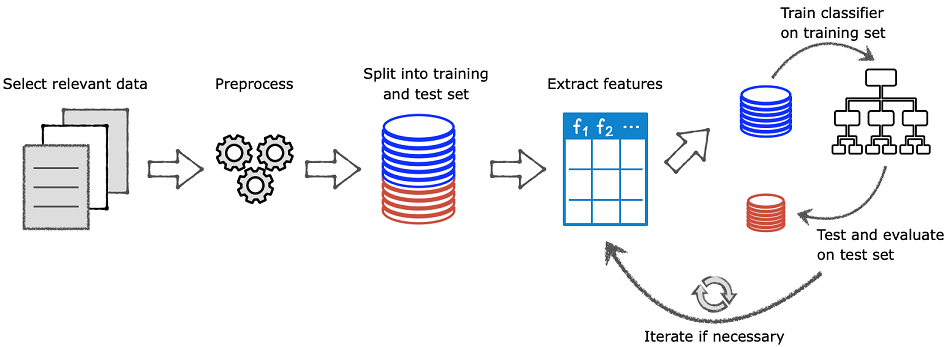
Email filtering, popularly known as “spam classification,” is one of the earliest examples of automatic text classification, which impacts our lives to this day.

A





|  |
| --- |
|  |





Naïve Bayes

 its content. This type of probability, when the outcome (class of “spam” or “ham”) depends on the condition (words used as features), is called conditional probability. For spam detection, you estimate P(spam | email content) and P(ham | email content), or generally P(outcome | (given) condition)

|  |
| --- |
|  |

