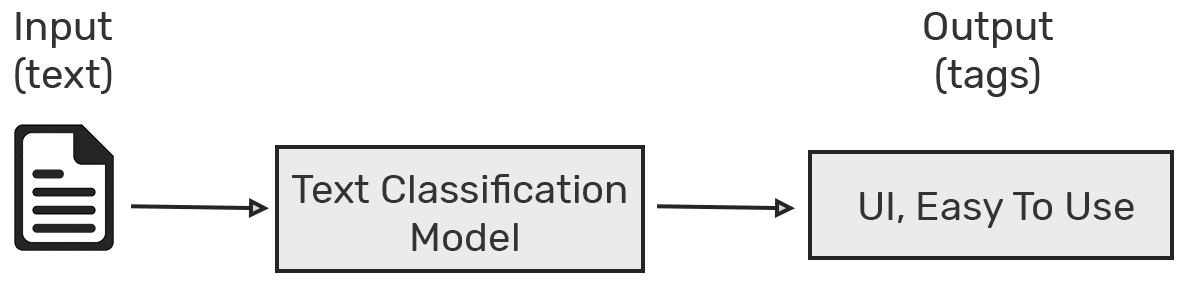
Text Classification

Text classification is the process of assigning tags or categories to text according to its content. It’s one of the fundamental tasks in [Natural Language Processing](https://monkeylearn.com/blog/definitive-guide-natural-language-processing/) (NLP) with broad applications such as sentiment analysis, topic labeling, spam detection, and intent detection.

A classifier can take this text as an input, analyze its content, and then and automatically assign relevant tags, such as UI and Easy To Use that represent this text:



## **How Does Text Classification Work?**

There are many approaches to automatic text classification, which can be grouped into three different types of systems:

* Rule-based systems
* Machine Learning based systems
* Hybrid systems

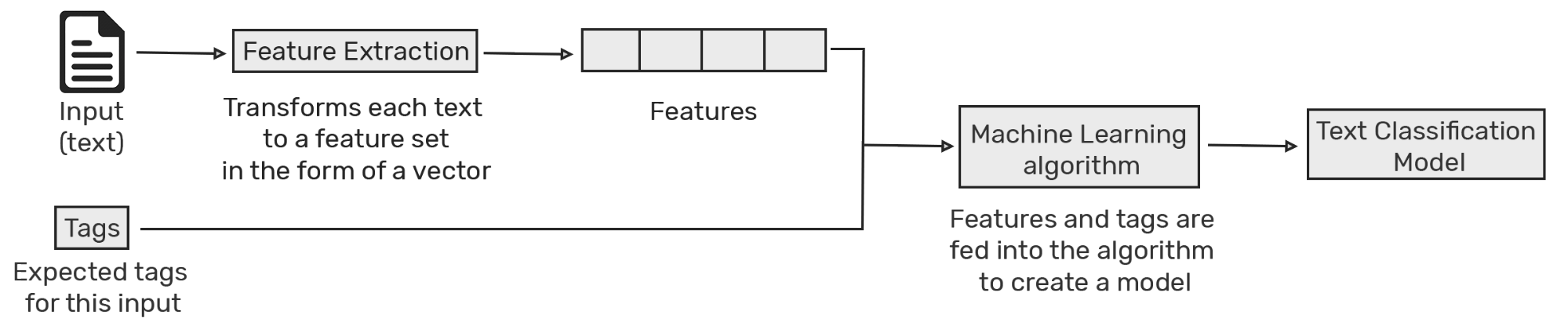
### **Machine Learning Based Systems**

Instead of relying on manually crafted rules, text classification with machine learning learns to make classifications based on past observations. By using pre-labeled examples as training data, a machine learning algorithm can learn the different associations between pieces of text and that a particular output (i.e. tags) is expected for a particular input (i.e. text).

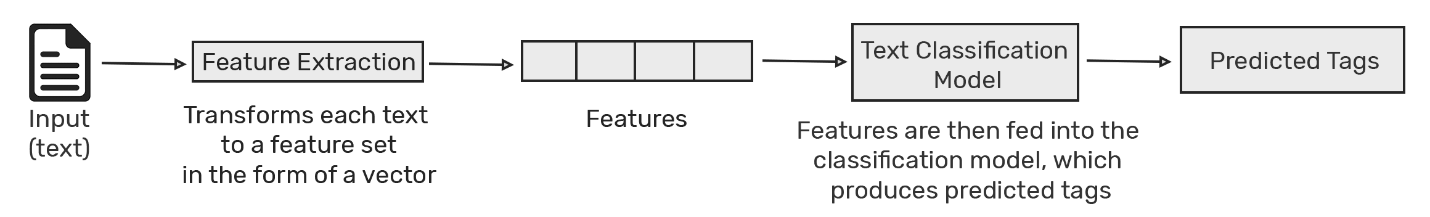
The first step towards training a classifier with machine learning is feature extraction: a method is used to [transform each text into a numerical representation](https://monkeylearn.com/blog/beginners-guide-text-vectorization/) in the form of a vector. One of the most frequently used approaches is [bag of words](https://machinelearningmastery.com/gentle-introduction-bag-words-model/), where a vector represents the frequency of a word in a predefined dictionary of words.

For example, if we have defined our dictionary to have the following words {This, is, the, not, awesome, bad, basketball}, and we wanted to vectorize the text “This is awesome”, we would have the following vector representation of that text: (1, 1, 0, 0, 1, 0, 0).

Then, the machine learning algorithm is fed with training data that consists of pairs of feature sets (vectors for each text example) and tags (e.g. sports, politics) to produce a classification model:



Once it’s trained with enough training samples, the machine learning model can begin to make accurate predictions. The same feature extractor is used to transform unseen text to feature sets which can be fed into the classification model to get predictions on tags (e.g. sports, politics):



Text classification with machine learning is usually much more accurate than human-crafted rule systems, especially on complex classification tasks. Also, classifiers with machine learning are easier to maintain and you can always tag new examples to learn new tasks.

#### **Text Classification Algorithms**

* Naïve Bayes
* Support Vector Machines
* Deep Learning

##### **Deep Learning**

[Deep learning](https://medium.com/dair-ai/deep-learning-for-nlp-an-overview-of-recent-trends-d0d8f40a776d) is a set of algorithms and techniques inspired by how the human brain works. Text classification has benefited from the recent resurgence of deep learning architectures due to their potential to reach high accuracy with less need of engineered features. The two main deep learning architectures used in text classification are [Convolutional Neural Networks](https://machinelearningmastery.com/crash-course-convolutional-neural-networks/) (CNN) and [Recurrent Neural Networks](https://machinelearningmastery.com/crash-course-recurrent-neural-networks-deep-learning/) (RNN).