

[LOGIN](#)

[REGISTER](#)

Search for...

[Home](#)

[Technical Topics](#)

[Business Topics](#)

[Sector Topics](#)

ADVERTISEMENT

ADVERTISEMENT

[Home](#) » [Technical Topics](#) » [Machine Learning](#)

ADVER

Data Labeling for Machine Learning Models



cogitotech | September 7, 2021 at 6:00 am

Related Content



30 Features that Dramatically Improve LLM Performance – Part 3

Vincent Granville |

September 21,



Machine learning models make use of training datasets for predictions. And, thus labeled data is an important component for making the machines learning and interpret information. A variety of different data are prepared. They are identified and marked with labels, also often as tags, in the form of images, videos, audio, and text elements. Defining these labels and categorization tags generally includes human-powered effort.

Machine learning models which fall under the categories of supervised and unsupervised, pick the datasets and make use of the information as per ML algorithms. **Data labeling for machine learning** or training data preparation encompasses tasks such as data tagging, categorization, labeling, **model-assisted labeling**, and annotation.

Machine Learning Model Training

The majority of effective machine learning models use supervised learning, which uses an algorithm to translate input into output. Machine learning (ML) industries, such as facial recognition, autonomous driving, drones, and require



Quantum leaps in machine learning: Pushing the boundaries of AI capabilities

**Shafeeq
Rahaman** |

September 9,
2024 at 12:40
pm



30 Features that Dramatically Improve LLM Performance – Part 2

**Vincent
Granville** |

September 1,
2024 at 12:46
am



How will machine learning transform HR strategies?

Pritesh Patel |

August 22, 2024
at 2:24 pm

supervised learning. And as a reason their reliability on the labeled data increases. In supervised learning, sometimes, machine learning models can also work to predict loss reduction. This instance is referred to as empirical risk minimization. For preventing such scenarios, data labeling and quality assurance must be vigorous.

In machine learning, as a norm, there are three main types of data sets that are utilized – dimensionality, sparsity, and resolution. And the data structure can also vary depending on the business problem. Textual data can be based on records, graphs, and order, etc. The human-in-the-loop uses labels to identify and mark predefined characteristics in the data. If the ML model requires to predict accurate results and also develop a suitable model, the dataset quality must be maintained. For example, labels in a data set identify whether the image has objects like a cat or a human, and also pinpoint the shape of the object. In a process known as “model training,” the machine learning model employs human-provided labels to understand the underlying patterns. As a result, you’ll have a trained model that you can use to generate predictions and develop a customized model based on fresh data.

Use Cases of Data Labeling in Machine Learning

Several use cases and AI tasks pertaining to computer vision, natural language processing, and speech recognition, computational instances need appropriate forms of data labeling.

1. Computer Vision: To produce your training dataset for a computer vision system, you must first label images, pixels, or key spots, or create a bounding box that completely encloses a





digital image. Once the annotation is done, a training data set is produced and the ML model is trained depending on it.



2. Natural Language Processing: To create your training dataset for natural language processing, you must first manually pick key portions of text or tag the text with particular labels. Tag and justify labels in the text for the training dataset. Sentiment analysis, entity name identification, and optical character recognition or OCR are all done using natural language processing approaches.

3. Audio Annotation: Audio annotations are used for machine learning models which use sounds in a structured format for example – extraction of audio data and tags. NLP approaches are then applied to tagged sounds to interpret and obtain the learning data.

Maintaining Data Quality and Accuracy in Data Labeling

Normally, the training data is divided into three forms – training set, validation set, and testing set. All three forms are crucial for learning the model. Gathering the data is an important step to collating raw data and properly defining the attributes, in order to get them labeled.

Machine learning datasets must be accurate and of high quality. Accuracy refers to how accurate each piece of data's labeling is in comparison to the business problem and what it aims to solve. Equally crucial are the tools which are used for labeling or annotation of data. [AI platform data labeling services](#) form the core for developing dependable ML models for artificial intelligence-based programs.

Cogito is one of the best data labeling companies, which offers quality training data for the machine learning industry. It makes use of labelbox model-assisted labeling,

The company has set the industry standard for quality and on-time delivery of AI and ML training data by partnering with world-class organizations. Cogito is well known in the AI community for providing reliable datasets for various AI models as the company fully supports data protection and privacy legislation. Cogito provides the clients with complete data protection rights that are governed by the norms and regulations of a [GDPR and CCPA](#), ensuring total data privacy.

[Machine Learning](#)

◀ PREVIOUS

[The Growing Impact of AI in Financial Services: Six Examples](#)

NEXT ▶

[Big Data Analytics: The Role it Plays in the Banking and Finance Sector](#)

Leave a Reply

Your email address will not be published. Required fields are marked *

Name *

Email *

Website

Comment *

☐

Save my name, email, and website in this browser
for the next time I comment.

Post Comment

[About Us](#)

[Contact Us](#)

[Partner with Us](#)

[Advertise with Us](#)

[Write for Us](#)

[RSS](#)

[Legal](#)

[Terms of Service](#)

[Privacy Policy](#)

[Do Not Sell or Share My Personal Information](#)

[Cookie Preferences](#)

© 2024 TechTarget, Inc.