

# A Comprehensive Guide to Data Labeling

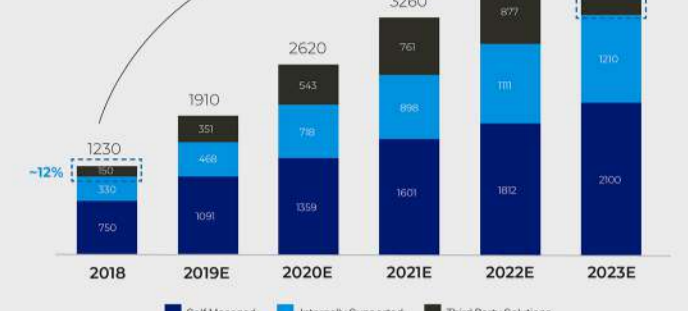
## What is Data Labeling?

Data labeling is simply the process of labeling or tagging relevant information/metadata in a dataset to let machines understand what they are. The dataset could be in any form i.e., image, an audio file, video footage, or even text.

## Global Data Labeling Market

Global data labeling spend on third-party solutions is estimated to be >7X in 2023 as compared to 2018

Global Data Annotation Spend (in \$ mn)



Refer: NASSCOM Data Annotation Report Page 09

## 7 Data Labeling Challenges

- Approx. 80% of the time is spent on Data Preparation
- Struggling to handle a vast workforce
- Limited access to cutting-edge tools & technologies
- Controlling the cost of Data labeling
- Falling short on consistent & quality data tagging
- Scalability
- Fail to comply with data security guidelines

## Types of Data Labeling



### Audio Classification

- Audio Collection
- Audio Segmentation
- Audio Transcription



### Image Labeling

- Image Classification
- Image Segmentation
- Image Key point Annotation
- Image Collection



### Text Labeling

- Text Classification
- Text Extraction



### Video Labeling

- Video Segmentation
- Video Classification
- Video Collection



### 3D Labeling

- 3D Object Tracking
- 3D Segmentation

## 4 Key Steps in Data Labeling



## Factors to consider while choosing the right tool

- » Defining Your Use Case
- » Expert Annotators
- » Establishing Quality Standards
- » Compliance Complexities
- » Open Source & Freeware
- » Budget

## Industries

- Medical AI
- Autonomous Vehicle
- Finance
- Retail
- Technology
- Geospatial

## Build Vs Buy

Build	Buy
<b>Pros:</b> <ul style="list-style-type: none"> <li>• Full control over the entire process</li> <li>• Faster response time</li> </ul>	<b>Pros:</b> <ul style="list-style-type: none"> <li>• Faster time-to-market for first movers advantage</li> <li>• Access to the latest in tech in line with industry best practices</li> </ul>
<b>Cons:</b> <ul style="list-style-type: none"> <li>• Slow and steady process. Requires patience, time, and money.</li> <li>• Ongoing maintenance and platform enhancement expenses</li> </ul>	<b>Cons:</b> <ul style="list-style-type: none"> <li>• Existing vendor offering may need customization to support your use case</li> <li>• The platform may support requirements &amp; doesn't assure future support.</li> </ul>