**What is the difference between labelled and unlabelled data? (Write in your own words)**

**Label Data**: when we have collected data in any form like Audio, video, images, graphs, tabular and tensor. We have to annotate or label this data. Label data is always annotated by humans (also called supervisar). Classification, Regression are common types to handle label data.

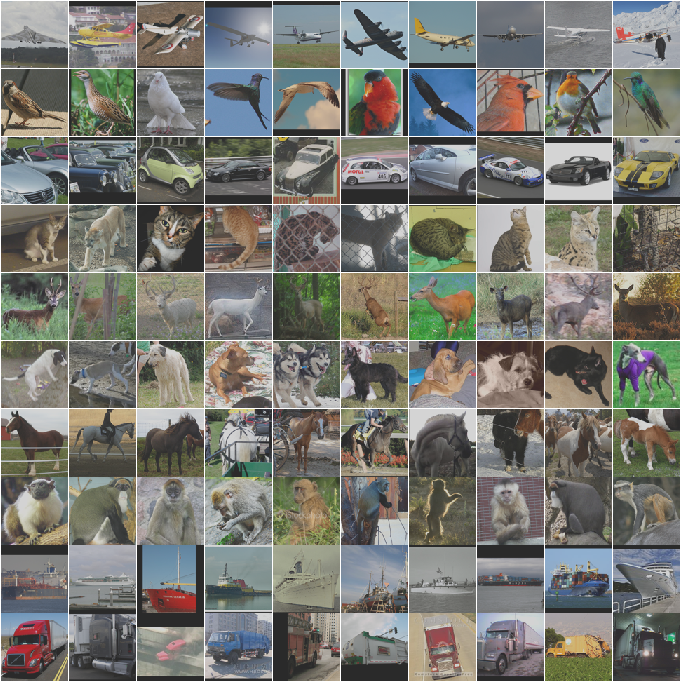
Examples:

|  |  |
| --- | --- |
| Data | Label |
| (Image, video, data) | cat |
| (sound data) | Qasim |
| Amount , Age, Login year  50,000 , 25 , 2010 | Purchase  Yes |

**Unlabel Data:** When you have collected data in any form like, Tabular data, Image Data, Video Data, Audio in Raw form. Those are not annotated or label by Supervisor or Human is called unlabeled data. We also handle these types of data with clustering or dimensionality reduction.

Examples:

|  |  |
| --- | --- |
| Data | Label X (If not available) |
| (Image, video, data) | X |
| (sound data) | X |
| Amount , Age, Login year  50,000 , 25 , 2010 | X |



**A Proposal for Local k Values for k-Nearest Neighbor Rule**

By Nicolás García-Pedrajas et al (2017)

**Summary:**

According to the above paper KNN algorithm is also most famous and most widely used by ML practitioners. It has interesting features like generalization, easy to implement. This simple technique can significantly outperform the standard k-NN rule for both standard and class-imbalanced problems in a large set of different problems.

In computer vision, it has achieved good performance, such as face recognition , articulated pose estimation, and character recognition.

They have found new algorithm for finding local K value for k-NN classifier . It's performed well on both data set standard dataset or Imbalance class datasets. They compared with standard approach to finding K value using tenkfold CV and the four different version of K-NN