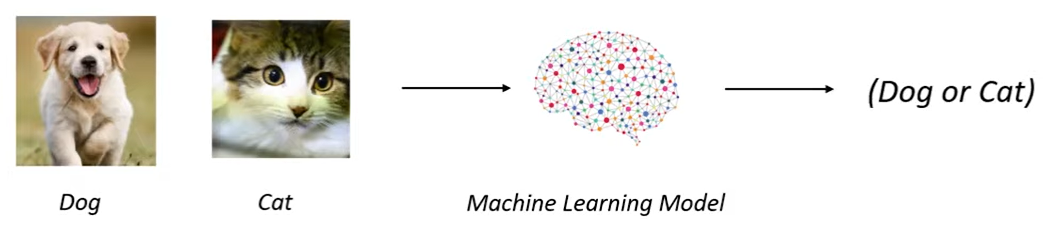
**Module 4: Data Collection & Processing**

* This module is on Data collection and Data Processing.
* In this module we will learn where we can get data for our Machine learning projects and how you can get this data.

**Agenda:**

1. Importance of Data in Machine Learning.
2. Where to collect the data.
3. Demonstration of data collection.



**Data for Machine Learning:**

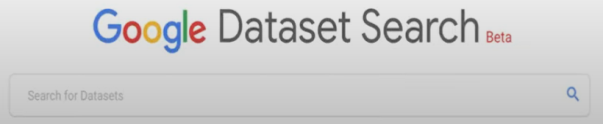
* We have the images of cat & Dog we want a system to see this image and recognize weather the image represents a dog or a cat.
* This is the problem statement we have to solve.
* What we can do is we can feed the images of dog & cat to a machine learning model. It can be a neural network (or) support vector machine model (or) any other model.
* Now this model will find the patterns in those images. So let’s say for an example if eyes are small it may represents cat and if the size of the image is small it may represents a cat. So on..
* All those features will be recognized by the model.

This is how you use data in machine learning.

* Here we are not talking about 10 or 20 images of cats or dogs.
* We often deal with thousand (or) even lakhs of data.
* We may deal with thousand or lakhs of images of dogs and also thousand or lakhs of images of cats.

**Where to collect the Data?**

The most important website where we can collect the data is

1. Kaggle
2. UCI Machine Learning Repository.
3. Google Dataset Search

**Kaggle:**

* Among All kaggle is the most important website where we can collect the data.
* Kaggle also host competition for the Data Science and Machine Learning so you can participate in those competition and you need to submit the code and it will be evaluated and get the prize.

**UCI machine learning repository:**

* In UCI machine learning repository you can collect data you want for doing machine learning projects.
* This is the logo of UCI machine learning repository.

These two websites you have so many number of datasets which you can use in doing machine learning projects.

**Google Dataset Search:**

* This is very similar to Google search and this is exclusively for searching datasets.
* In these three sites you can find datasets.
* You can also make dataset for yourself for example if you want to make face recognition system which detects your face.

In this we will see how we can get your data from Kaggle or UCI machine learning repository or Google dataset search.

**Hands on data collection.**

Search for google dataset search

Official site -> dataset.search.research.google.com

In this search for -> boston house price -> where you can find many sites offering this dataset.