

Jaime Andrés Montemayor Molina  
A01176573

### Tarea 3

1- What are the 3 methods for data acquisition? Provide a brief description of each one and a data collection technique for each method.

Data discovery: Is researching already published data.

Data Augmentation: Is adding the data from a dataset to another.

Data generation: Is creating your own datasets.

2- Enumerate 3 platforms for sharing datasets and 2 for searching datasets.

**Sharing:**

DataHub

Google Fusion Tables

Data Market

**Searching:**

Data Lake

Dataraman

3- What is data augmentation? Describe 1 way to augment a dataset.

It is the addition of data to an existing dataset., one way to augment a dataset is to derive latent semantics from data.

4- Accurately labeled data ensures that ML systems establish reliable models for pattern recognition, which forms the foundations of every AI project. Describe what are the disadvantages of doing manual labeling and describe 2 alternative approaches to label data.

Disadvantages of doing manual labeling are: the limited amount of labour that can be made in a certain amount of time and the cost of labour

2 alternative approaches: Using existing labels and Crowdsourcing.

5- Mention 3 crowd-based services you can use to label data.

Weak labels

Machine learning task

Data type

6- In machine learning, what is human in the loop?

It is allowing humans to interact with a system that they have and use the results to solve the cases for prediction with a low level of confidence. Using the feedback, the system adjusts the scope of its answers.

7- What is data cleaning? Enumerate 2 cleaning system/framework.

Data cleaning is fixing or elimination of corrupt, duplicate, false or incomplete data from a dataset.

Data cleaning systems: HoloClean / ActiveClean

8- Explain transfer learning.

Transfer Learning is the use of an old model as a starting point for the new model. It is basically using the knowledge of a previous machine on a new one.

9- Explain a concept that you found interesting and was not included in the previous questions.

The Fact Extraction concept was interesting, and I thought of the almost infinite uses it has. Considering that it can take anything from the web for the entity requested.