

Data Labeling System

Requirements Analysis Document (RAD)

Purpose of the system

Data labeling, in the context of machine learning, is the process of detecting and tagging data samples. The process can be manual but is usually performed or assisted by software. Data labeling is an important part of data preprocessing for ML, particularly for supervised learning, in which both input and output data are labeled for classification to provide a learning basis for future data processing. This system can be used to label news from on online newspaper, articles as sports economy ,world ,politics other also can be used to label customer comments in e-commerce web sites.

Stakeholders

- Murat Can Ganiz
- Lokman Altın

Scope of the system

Data labeling service comprises many different tasks. This project we use data labelling system for users comments. Data labelling system can be used by customer services, complaint/request sites and in extracting data from these databases.

Functional Requirements

- System will be a multi-user system.
- There must be many to many relationship users and instances.
- Program will be getting user information as a json file.
- Program must support easily pluggable labeling mechanisms.
- The labeling mechanism will be random labeling mechanism.
- Program will be write results in a json file

Non-Functional Requirements

- A user can label many instances
- An instance can be labeled by one or more users (possibly with different class labels).
- The LabelingMechanism needs to assign one (or more if multilabeled) of these labels to a given instance.

Glossary of Terms

User – People who label the given datas

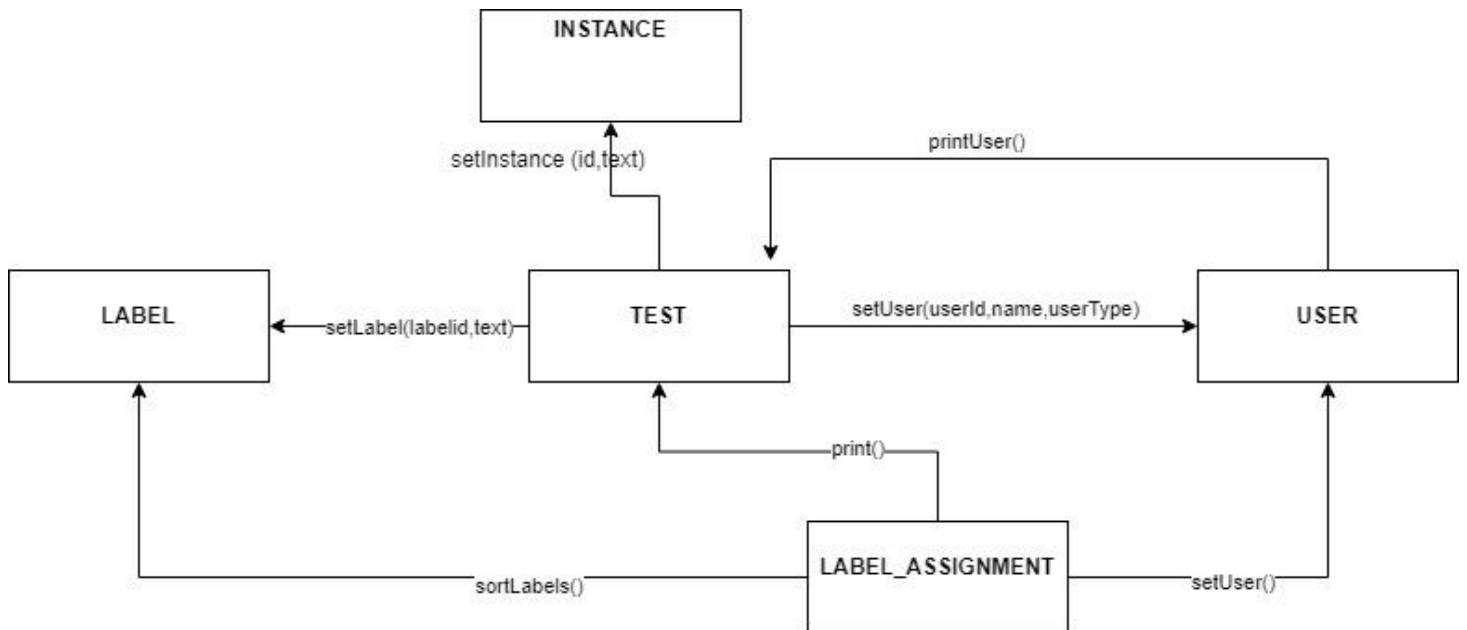
Label – A classifying phrase or name applied to a person or thing

Instance – Sentences that need to be labeled

Dataset – A collection of related sets of information

Labeling mechanism - Mechanism that we used in our project to label datas

Domain Model



CREATED BY GROUP:

- 150117034 Berra Mercan
- 150117510 Elif Gökpınar
- 150117042 Ezgi Doğruer
- 150117061 İsra Nur Alperen
- 150117057 Rümeyssa Öztürk
- 150116010 Şükriye Soyer
- 150114036 Ömer Kaya