

# DATA LABELING SYSTEM

## REQUIREMENTS ANALYSIS DOCUMENT

### VISION

In the first iteration, our vision for Data Labeling System is assigning predetermined labels to a dataset. Our system will be developed as a simulation that takes arguments of user configuration file and input file. Simulation runs until labeling mechanism is done. System has used by the bots, so our mechanism should be like randomization factor.

### PROBLEM STATEMENT

Main objective is developing solution for some classification problems related to labeling mechanisms in datasets such as sentiment classification problem.

### SCOPE

Our Data Labeling System will be simulation for the first iteration. This system will be getting a dataset as a Json file. There are some labeling mechanism for processing. In this iteration, labeling mechanism will be random label mechanism which randomly chose one of the labels from the set of labels and assigns it to the instance. After processing all instances, the system will create output Json file.

### SYSTEM CONSTRAINTS

Our system will run on any Java IDE as a console application with customized Json library original library was Json-Simple.

### STAKEHOLDERS

Murat Can Ganiz (Customer)

Lokman Altın (Customer)

Edanur Öztürk

Feyza Nur Bulgurcu

Sueda Bilen

Ömer Erkan

Zehra Kuru

Farouk Tijjani Mohammed Deribe

## GLOSSARY OF TERMS (ALPHABETICALLY LISTED)

Dataset: The structure that keeps labels and instances.

Instance: Content that need to be labeled in dataset and it must be labeled one time.

Json File: The file which includes meta data.

Label: Content that has used for categorizing the instances.

Labeling Mechanism: Method for labeling the instances.

Negative: Indicates instance content as negative.

Notr: Indicates instance content as notr.

Positive: Indicates instance content as positive.

Random Labeling Mechanism: The mechanism that label instances randomly

Simulation: Is an approximate imitation of the operation of a process or system.

User: who runs the system.

## USE CASES

- 1) User executes the Data Labeling System.
- 2) Simulation starts.
- 3) Dataset will be uploaded to system.
- 4) For each user, system decides labeling mechanism based on user's type
- 5) For each user, system randomly links some instances to users.
- 6) Random Labeling Mechanism will choose one of the labels from the dataset randomly and assigns it to an instance.
- 7) System creates a log for the labeling processes.
- 8) When there is no remaining possibilities system stop the simulation
- 9) System creates output file based on assignments.

