# REQUIREMENT ANALYSIS DOCUMENT

## Vision:

Data labeling is a process of classfining instances(text,images,recordes) and assigning each group of instances to a suitable label - category - using a classification model chosen by expert users. This process gave us a result dataset which usually helps in training machine learning models which use these datasets to understand the pattern and take a decision based on the predicted pattern.

## Scope:

The data labeling project will provide :

- Multi-users will be able to use the program and choose a labeling mechanism .
- the program will support the random labeling mechanism only.
- In this iteration the user will be able to label the same instance more than once and many users can label the same instance .
- Beside the bot user we provide in the last two iterations in this iteration real users will be allowed to interact with the program via command line / terminal .
- Many users are allowed and all of them stored in the config file .
- Human users will enter their name and password and the program will check if they are already stored in the config file, if not the program will ask the user to enter the information again .
- If the username and password are empty then our program will act like the last two iterations and will use bot users .
- We provided a new labeling mechanism called SentimentLabeling beside the randomLabeling .
- The final instance label will be the most frequent class label, if all labels have equal frequency the program will choose one randomly .
- The program will provide User Performance Metrics and Instance Performance Metrics and Dataset Performance Metrics.
- Those three matrices will be used to detect the changes in user labeling behavior and the different labels that the same instance could have .
- Those matrices will give us an idea about the quality of the data labeling and the quality of the users.
- After calculating the matrices end the program will print the result report as text in the command line screen and json file.
- After the classification process is done the program will create a json file containing all the instances with suitable labels and user information .

## System constraints

GUI part is not included at this iteration . The required displayments are made in the console.

## Project Members:

- 150117046 Ahmet TURGUT
- 150117048  Fatih BAŞ
- 150117010 Süleyman Ahmet SÖNMEZ
- 150117035 Alperen KÖKER
- 150117037 Koray Emre ŞENEL
- 150118504 Mehmet Etka UZUN
- 150119529 Esraa Ismail Taha TURKY

## Glossary of Terms:

- Labeling mechanisms : models process the date and assign instances to suitable labels
- Random labeling  : model assign random label form the list of labels to an instance
- Multi-labeling : proved more than one label to one instance
- User Performance Metrics: matrices track the behavior of the user at the classification process.
- Instance Performance Metrics: matrices track the different labels that are assigned to the same instance .
- Dataset Performance Metrics: matrices track the changes in the dataset during the classification process.
- SentimentLabeling : model for instance labeling based on the number of positive or negative words in the instance .