Dear AGT Data Science applicant,

The task at hand is a multi-class classification problem, for which both a training and a test (validation) set are provided as csv files, 'train.csv' and 'test.csv' accordingly.

What we ask is that you work on this classification task by building a classifier using only the training data, with the goal of achieving the best performance possible on the test data, classifying as correctly as possible the 'label' variable.

You can use one out of the following 3 programming languages: Python, R or Matlab.

To show your work, we expect to receive a short writeup with your observations about the problem, your thought process throughout your work with it, any interesting findings you have and, finally, a descriptions of the results you have obtained. This report should be in no case longer than 2 pages. We would also like to get any code you write for this project, and would appreciate it if it was informally commented so that its easier for us to read.

In case you want to combine code and the write up, feel free to use Jupyter Notebook (for Pyhton) or RMarkdown (for R) to write the code in an interactive computational environment.

Best of luck,
The AGT Data Science Team