.csv format dataset is taken from www.kaggle.com

That dataset is raw, which simply includes mixed rows and non-ascii inprocessable entries.

A Java application that arranges row order, fixes mixed and separated rows and loops for removing non-ascii characters has been created.

Java is ***optional***, made it for the sake of easiness but Python possible way faster and easy-to-implement.

Columns, indicating features that will not be used in the training and testing, are removed.

Just assign all cells to ***empty***.

Put this version of data to matrix called ***raw***

Use a toolbox, that is trained by some part of (I used half) dataset and some images.

Now got 3 output, the one named result keeps the success of the algorithm.

Get the error distribution from that struct. Put this to the regression matrix

Put the numbers in the preferences matrix

Get Hex color info of users, convert it to CIE-Lab color space to find meaningful distances of preferences

Get the error distribution from that struct. Put this to the regression matrix

Perform multinomial logistic regression by using randomly selected half of the data.

Remaining half is tested.