The following python code is used to compare two different or similar items which are entered by the user. The input from the user should contain the following in the same order:

1. Name of the item
2. Color of the item
3. Density of the item.

In the training\_data.csv file we have the sets of training data which I have created in order to train the algorithm. Here the data contains the Index, Density and the class it belongs to, based on the density of the entered item.

The density ranges from 0-100 in my dataset.

The density ranges from:

1. 0-10 which falls in class A (represented as 0)
2. 11-20 which falls in class B (represented as 1)
3. 21-30 which falls in class C (represented as 2)
4. 31-40 which falls in class D (represented as 3)
5. 41-50 which falls in class E (represented as 4)
6. 51-60 which falls in class F (represented as 5)
7. 61-70 which falls in class G (represented as 6)
8. 71-80 which falls in class H (represented as 7)
9. 81-90 which falls in class I (represented as 8)
10. 91-100 which falls in class J (represented as 9)
11. 100+ which falls in class J (represented as 9)

The items if similar will fall in the same cluster. But this is not enough to say whether the two items are different or same. So after clustering the items, we need to compare the colors of the items. If the colors are also same then we go by the label of the item, if they match as well then the items are similar.

The algorithm is compatible with any other dataset as well with slight modifications to the code.

Since any product can be of any color we don’t specify colors in the dataset. Instead we compare the two sets of inputs by the user and check for the color, if found same, then the items are same otherwise different.