



PLAN OF WORK

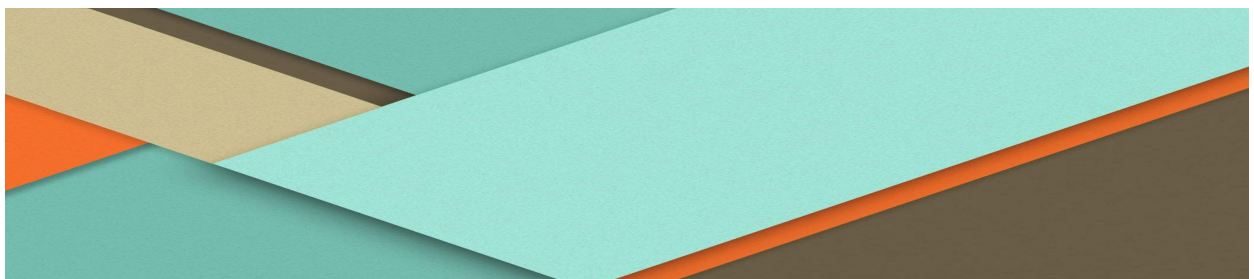
Alvaro Walter Bayardi Mederos
Montevideo, Uruguay, South America

13/06/2022

Course of Coursera:

Foundations of Data Science:
K-Means Clustering in Python

(by Universidad de Londres & Goldsmiths, University of London)



Overview

Through a course project, a report will be simulated for a bank that is considering automating the detection of counterfeit banknotes.

Objectives

1. Make the report well written and include all the required information.
2. 2 to 4 pages

Specifications

Your report must contain the following information:

- the purpose of the Data Science project;
- data description;
- methods: how the data was analyzed;
- summary of results;
- recommendations for your client.

(Use the descriptions and graphics you produced for previous assignments.)

Purpose of This Project:

The purpose of this data science project is to demonstrate to a bank about automation for counterfeit bill detection.

Bank staff do not need to be data science experts, but they do need to understand how it works and how effective it is.

The analysis is based on V1 and v2 values in the Banknote Authentication dataset.

Description of the data;

The complete set of data on the distinction between authentic and counterfeit banknotes is available in OpenML:

<https://www.openml.org/search?type=data&sort=runs&id=1462&status=active>

The file in CSV format : 'Banknote-authentication-dataset.csv'

The data was extracted from images that were taken of genuine and counterfeit banknote-like specimens.

For digitization, an industrial camera generally used for print inspection was used.

The final images are 400 x 400 pixels.

Due to the object lens and the distance to the investigated object, grayscale images with a resolution of approximately 660 dpi were obtained.

A wavelet transformation tool was used to extract features from these images.

The data contains two distinct value types involving approximately 1300 numeric values in each.

```
In [8]: data.describe()
```

Out[8]:

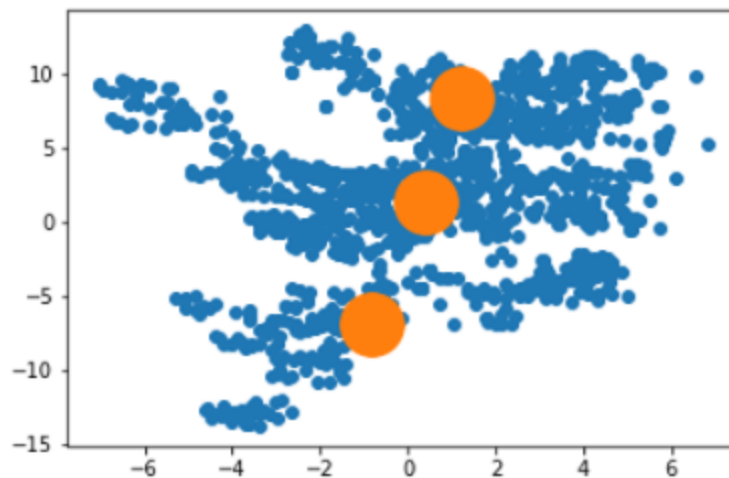
	V1	V2
count	1372.000000	1372.000000
mean	0.433735	1.922353
std	2.842763	5.869047
min	-7.042100	-13.773100
25%	-1.773000	-1.708200
50%	0.496180	2.319650
75%	2.821475	6.814625
max	6.824800	12.951600

```
In [10]: data.describe().sum()
```

Out[10]: V1 1376.603853
V2 1386.395975
dtype: float64

Method

The data was analyzed through K-Means Clustering in Python. K-Means Clustering is a concept that falls under Unsupervised Learning. This algorithm can be used to find groups within unlabeled data.



Result

In this case, the information that can be provided by the database of real banknotes and counterfeit banknotes is important to automate the detection of counterfeit banknotes already in circulation and also the detection of suspected new counterfeit banknotes.

Recommendation

In order to use this tool with better results, a more complete data set is needed. Preferably a number that is not too small.