Detection of Counterfeit Banknotes using K-Means Clustering

Marina Rivosecchi

Date: 23/04/22

Objective

Bank ABC considers automating detection of forged banknotes. The objective of this report is to demonstrate that with an adequately performing model, instances of forged bank notes can be reliably classified.

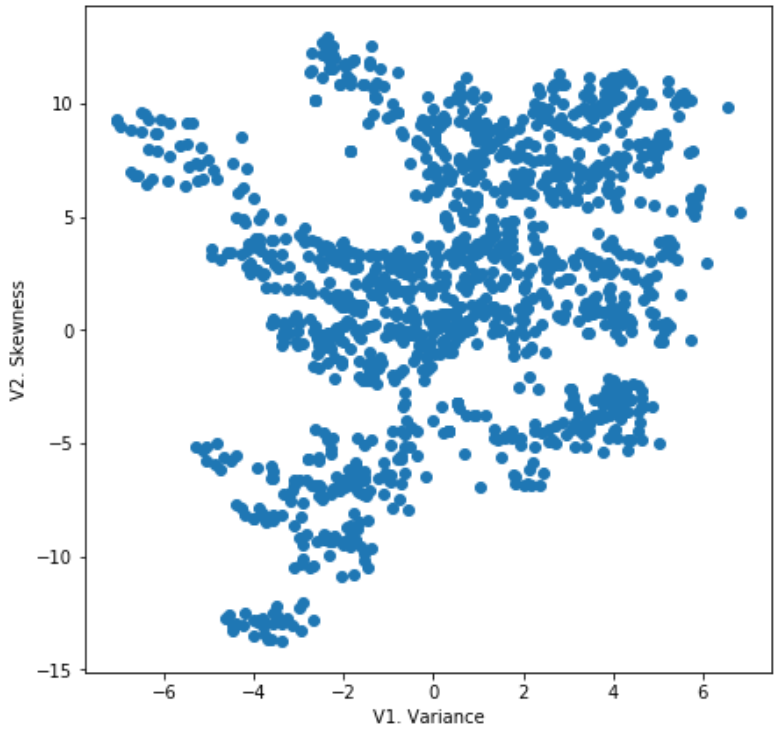
Description of the data

Data was extracted from images that were taken from authentic and counterfeit banknote-like specimens. For digitization, a digital camera typically used for print inspection was used. A Wavelet Transform tool was used to extract features from these images *(Source: OpenML)*.

The features (or attributes, variables) in the dataset are:

1. V1. variance of Wavelet Transformed image (continuous measure)
2. V2. skewness of Wavelet Transformed image (continuous)
3. V3. curtosis of Wavelet Transformed image (continuous)
4. V4. entropy of image (continuous)
5. Class (target). Presumably 1 for genuine banknote and 2 for forged banknote

For the purposes of our initial data model, only the first two variables (V1 and V2) were used.



Next figure shows a sample of the dataset:

V1 **Variance of the transformed image**	**V2** **Skewness of the transformed image**
3,6216	8,6661
4,5459	8,1674
3,866	-2,6383
3,4566	9,5228

The statistical measures of the variables used in the model are:

V1. Variance of the transformed image	
Maximum value	6.8248
Minimum value	-7.0421
Mean	0.4337
Standard deviation	2.8417

V2. Skewness of the transformed image	
Maximum value	12.9516
Minimum value	-13.7731
Mean	1.9224
Standard deviation	5.86694

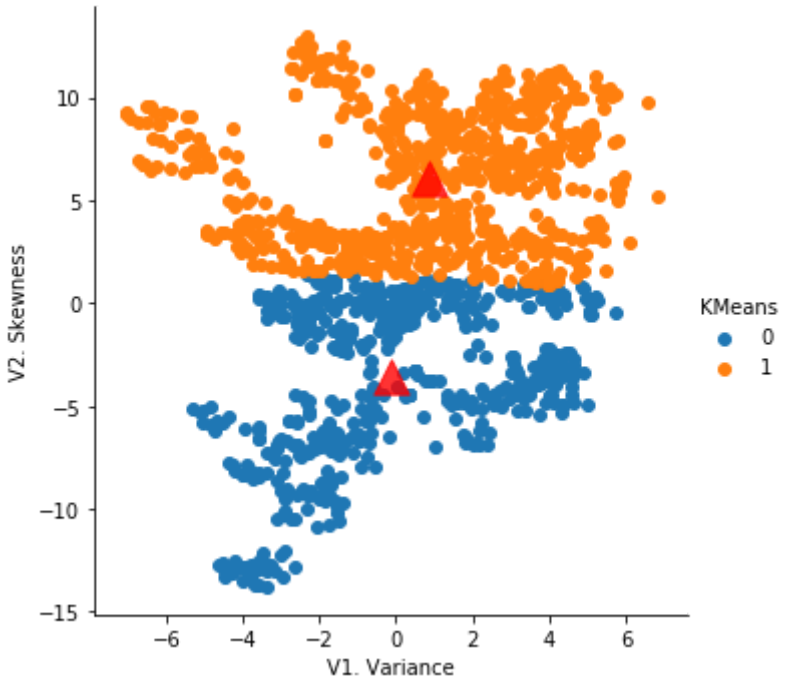
Limitations of the data :
- Our model uses only two attributes of the banknotes' images – variance and skewness.
- Limited size of the data set (1372 instances).
- Inherent inaccuracy of any data model.
- One of the identified clusters contains both fake and genuine banknotes to seemingly the same degree (so if the banknote is assigned to this cluster, its classification is likely to be either genuine or fake).

Approach

The model utilizes a k-means clustering algorithm, which is an unsupervised learning algorithm that is usually used to find groups which have not been explicitly labeled in the data. In this case, it is to identify banknotes as either genuine or fake with some degree of reliability.
According to the combination of banknotes' features (in the modelled case two: variance and skewness of the banknote transformed image), the model assigns the tested banknote to various clusters that suggests (again, with a different degree of reliability) classification of either genuineness or forgery.
The accuracy of the resulting model was compared to actual results – validated classification of banknotes which was also part of the data file.

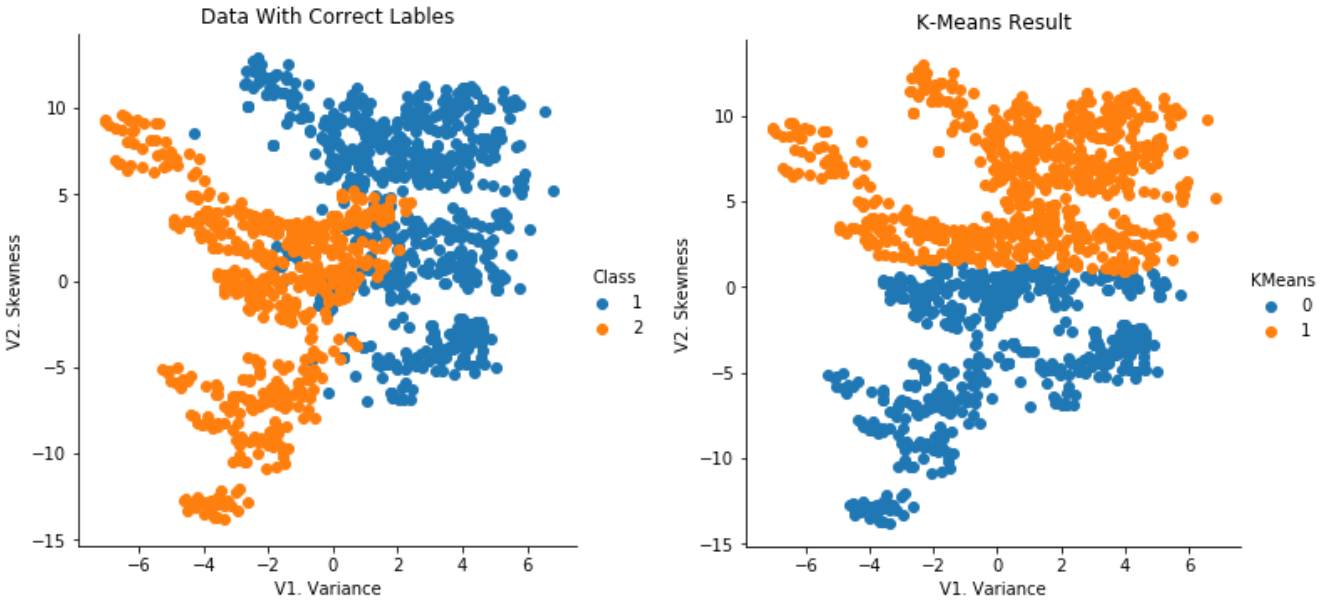
Summary of the results

In this project, the K Means algorithm is fitted to the data in order to predict whether the banknotes in our dataset are genuine or fake.



After running K-Means for 9 times, the results are very similar, which means the K-Means is stable.

To evaluate the clustering model, the original data with labels and our K-means analysis result is compared:



After comparing the value of each observation in both situations, it is concluded that the accuracy of this K-Means Model is 65.3%.

Recommendations

The accuracy of the analysis is 65.3%, which is considerably better than a random guess. Based on the data analysis performed, **the k-means clustering algorithm is recommended to help with fake banknotes recognition**.

It's also recommended to explore whether other combinations of features are able to build a better model with higher accuracy. In this regard, it is highly recommended to split the data into train, test and validation sets to examine the model's ability to classify new data as forged or genuine banknotes.

To complement the analysis in this project, it's advised to decompose the data into components to analyze what banknotes features are best explaining how to classify a banknote as real or fake. It's also advised to explore other algorithms in combination with model build in this project.