

## AUTOMATING THE DETECTION OF FORGED BANKNOTES

### Purpose of the project

The purpose of the project is to classify banknotes as **Forged** or **Genuine** depending on the parameters given in the dataset. These parameters are the characteristics that are derived from different samples of banknotes using the wavelet transforming tool measured using devices. An unsupervised learning metric, K-means clustering will be used to do cluster analysis on the dataset.

### Description of the Dataset

We are using a simplified version of the banknotes dataset available at OpenML. It contains 2 features — V1 (variance) and V2 (skewness) which are numerical values which represent the features of the notes. The dataset is a mix of values for both Forged and Genuine banknotes. Hence, we are using Kmeans cluster algorithm to filter out both the categories.

To have a more understanding of what our dataset looks like, we did a statistical analysis of the data using measures like mean, standard deviation, min and max as summarized in Table 1. This showed that the range between values in both features V1 and V2 is quite large. Hence there is a need to normalize the values in the dataset and bring it to a single range.

	V1	V2
count	1372.000000	1372.000000
mean	0.433735	1.922353
std	2.842763	5.869047
min	-7.042100	-13.773100
Max	6.824800	12.951600

**Table 1:** Statistical analysis of the dataset

### Method of Analysis

K-Means Cluster analysis is the metrics to use because it is suitable to the dataset we have since there are no missing values in our dataset. Since the range in the features of the dataset is large, the first step is to normalize the dataset and bring it to a single range.

The next step is to classify our dataset into groups that we want to see in the result, in this case (forged and genuine). Thus, our K-means clusters will be two.

## **Summary of the Results**

KMeans performs the clustering assigns each data point to the appropriate group 1(forged) or group 2(genuine). Figure 1 is the visual representation to compare the K-mean cluster results and the actual labels of the dataset. The red cluster belongs to the forged notes while the green cluster belongs to genuine notes. The blue dots are the cluster centres. These are values which represent the centre part of a cluster. This point is then used to generate predictions for a new data point.

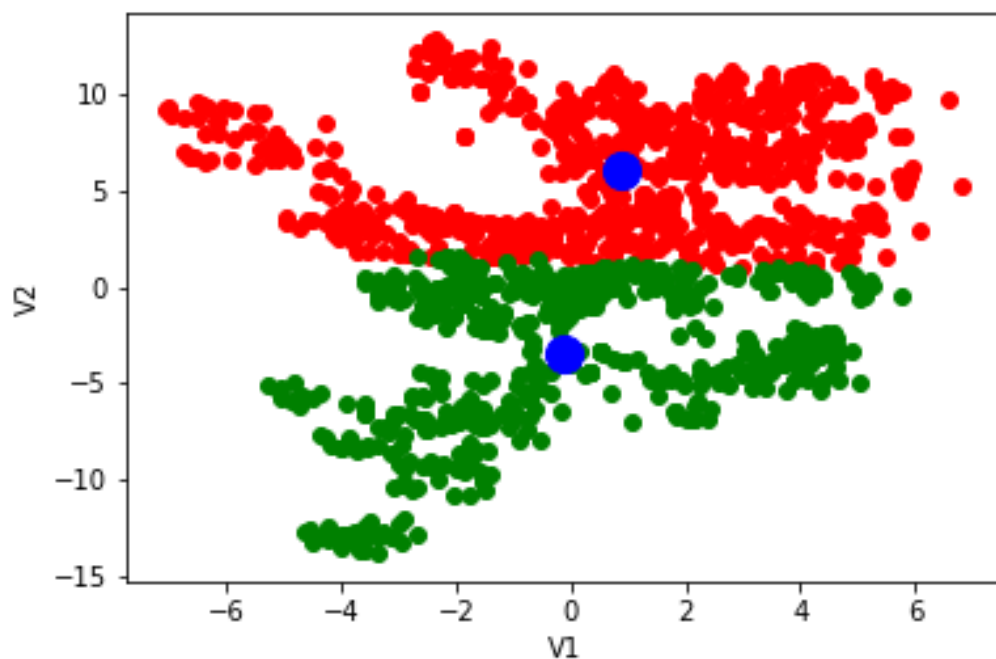


Figure 1: scatter plot visualizing the cluster analysis

## **Recommendations**

After careful analysis, it is worthy to note that the K-means cluster algorithm has proven effective in distinguishing between forged notes and genuine ones.

However, it is recommended that the bank have more information collected on the banknotes to better distinguish between genuine and forged banknotes and also increase the success rate of the process.