Automation Detection of Forged Banknotes

For Bank (Client)

Introduction:

According to the US Department of Treasury, there is an estimated $70 million in counterfeit banknotes currently in circulation. This represents a small yet significant portion of currency that is being used in transactions across the globe. Automating the detection of counterfeit banknotes can help identify malicious actors and save millions of dollars for your bank. With the use of advanced data science technology, we can assist in recognizing fraudulent notes.

Purpose of Project:

The purpose of this project is to help you (the bank) identify and classify counterfeit banknotes. Using the K-Means clustering algorithm, we can correctly classify which banknotes are authentic and which are counterfeit at a rate of 74%. Since the result of classification is binary (only two possible outcomes), the assumed rate of random guess is 50%. Our algorithm provides a 24% increase in the possibility of correctly identifying these notes.

Description of Data:

Looking at the data provided, we have two features, variance and skewness. This data was extracted from images that were taken from both genuine and forged banknotes. Using these images, a Wavelet Transform tool was used to extract these two features. We have 1372 independent images that are available in our sample. See *Figure 1* for basic descriptive statistics.

Figure 1:

V1: Variance	
Min	-7.0421
Max	6.8248
Mean	0.43374
Std Dev	2.84173

V2 skewness	
Min	-13.773
Max	12.9516
Mean	1.92235
Std Dev	5.86691

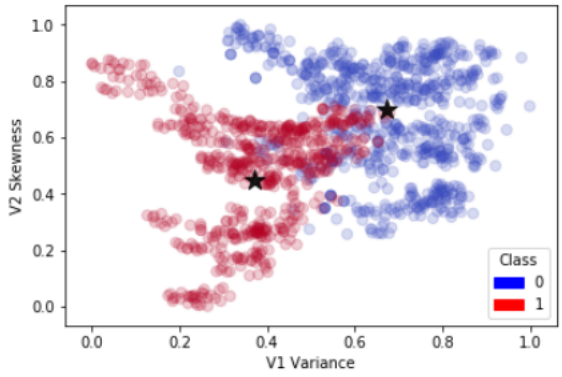
Methods:

Using the data provided, a K-Means clustering algorithm was used to identify two classes of banknotes. K-Means is an unsupervised machine learning algorithm that groups data points into clusters based on their similarity using a mathematical distance measure. The k represents the number of clusters, in this case we want to identify two clusters. This creates two distinct groups of data points that are comparable to each other, yet different to data points in the other groups.

Using the provided data, several iterations were run. We used the existing Class data from OpenML(Ground Truth) to check our model's accuracy in classifying the authentic or fraudulent banknotes. On initial runs, our model was only to yield a correlation of 29% when compared to the Ground Truth. Using 50% as a benchmark of random guessing, this model prediction performs significantly worse than random guess. After normalizing the data (organizing the dataset to a common scale), the model was increased to a correlation of 74%, a 45% increase and

24% improvement when compared to random guessing. This correlation means that our model's prediction ability is impactful and can be very useful in identifying genuine or forged banknotes.

Figure 2:



My Kmeans Prediction Model's correlation coefficient with Ground Truth from OpenML: 0.7416, or 74%.

In *Figure 2*, the graph shows our clustering results. The individual data points are colored blue or red based on the class they belong to according to the Ground Truth from OpenML. Our prediction model can successfully create two different clusters of data, as represented by the two black stars that are located centrally within each of the colored groups.

Summary and Recommendation:

Using our K-Means clustering model, we can predict if a given banknote is genuine or counterfeit with a 74% accuracy. If paired with different checks, we can help the bank identify and classify counterfeit banknotes and remove them from circulation with a high level of accuracy. Your bank should take our analytics backed methods to ensure a more rigorous level of authentication is used for transactions going forward.