Project Report:

Project Purpose:

For decades, the Banking sector has perpetually been faced with the challenge of counterfeit banknotes. In this project, I used machine learning to model a simple algorithm that can automate the process of detecting counterfeits from the real banknotes.

Dataset Description

Sourced from OpenML with Wavelet Transform used to extract special features from real and counterfeit banknote images, the data is contained in a csv file with 1372 rows 2 columns V1 and V2. Where:

V1 represents the Variances of the Wavelet Transform images, and

V2 represents the Skewness of the Wavelet Transform images.

(Note that n = 1372)

Analysis Method Description

In this project, I used K_means clustering as the machine learning algorithm to analyse the Banknote data. I repeated the clstering process by first creating 5 clusters, then 4 clusters and finaly settling on 2 clusters (to represent genuine or fake banknotes). Furthermore, I had to normalize my dataset for better and more visually desirable results.

Summary of Results (Findings):

Using the Numpy Python library, I was able to deduce the following from the dataset:

For V1 (Variances):

The mean is 0.43, the standard deviation is 2.84, and the range of the data is between -7.042 (min) and 6.825 (max).

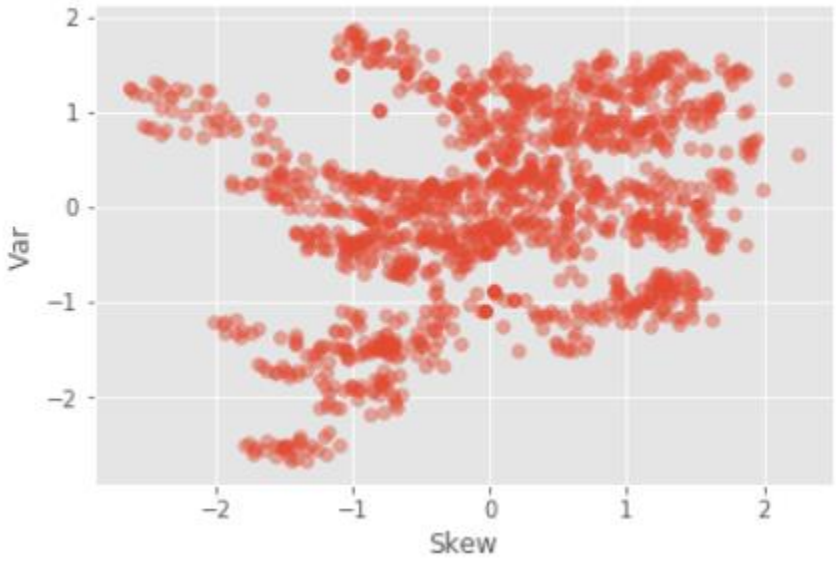
For V2 (Skewness):

The mean is 1.92, the standard deviation is 5.87, and the range of the data is between -13.773 (min) and 12.952 (max).

After the data was normed I was able to obtain the following scatter plots:

Before runing K-Means algorithm:

```
Out[65]: <matplotlib.collections.PathCollection at 0x7fa911b3ab00>
```

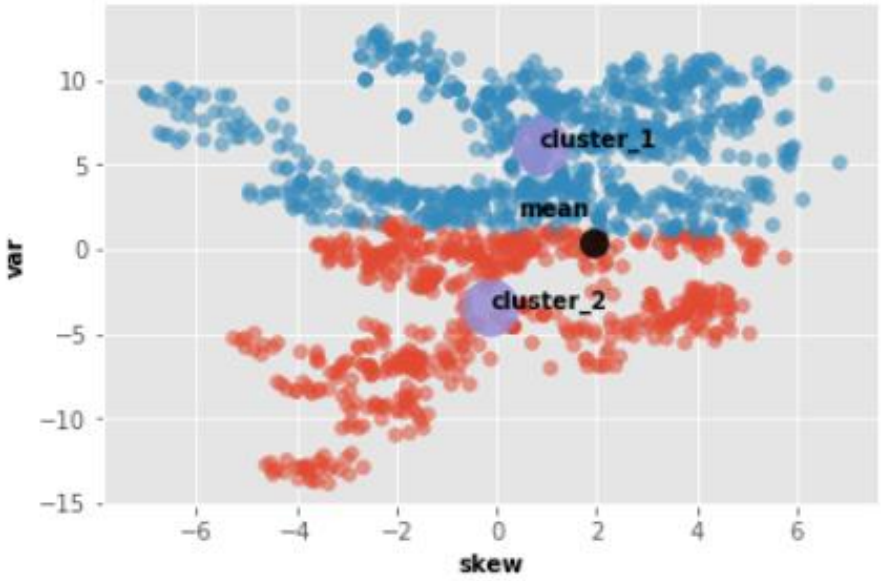


After Running K-means Algorithm using code:

```
km_Res = KMeans(n_clusters = 2).fit(Range_V1)

print([km_Res, Range_V1])
```

```
Out[97]: Text(-0.123767,-3.45591,'cluster_2')
```



NB :

To ensure stability of the results, I ran K-means 10 times.

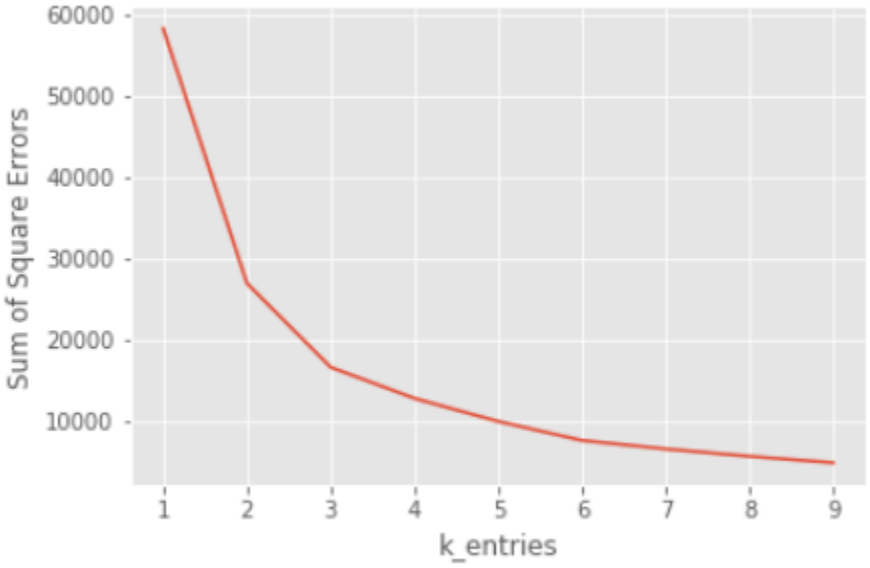
The algorithm achieved stability at cluster centers:

(0.86960048, 6.12717909) for Cluster 1, and
(-0.12376677, -3.45591265) for Cluster 2

The red data points as indicated above, clustered around Cluster 2, represents the genuine banknotes.
The blue data points as indicated above, clustered around Cluster 1, represents the genuine banknotes.

Elbow Plot

Lastly, I created the elbow plot to help further visualize the results:



Recommendations

Despite getting exemplary results from the K-means Machine Learning model, it is my opinion that the results could be improved. To provide room for improvement it is my recommendation that the number of samples be increased (especially the number of counterfeits) so as to furnish the algorithms with as much data as possible to chew and spit out the most desirable results.

Furthermore, to efficiently collect, analyze, and scan the banknotes (whether genuine or counterfeit) and enhance the analysis of the K-Means algorithm, the client should install a simple workflow whose wavelet transform technology is already predetermined.