Report on Automating the Detection of Forged Banknotes

Purpose of the Project

The main goal of this project is to create a system that can automatically detect fake banknotes. This will help the bank reduce losses from counterfeit money and improve overall security.

Description of the Data

The Banknote Authentication dataset contains 1372 instances with 2 features: variance('V1'), skewness(V2) of wavelet-transformed images of banknotes. there is no missing data .TO describe the data we calculated for both features (v1 vs v2): the mean value which is the average (0.433735, 1.922353), The standard deviation (2.842763, 5.869047) which measures how spread out the values are from the mean. A higher standard deviation means the values are more spread out, the minimum which is the smallest value in the dataset (-7.042100, -13.773100), and the maximum which is the largest value (6.824800, 12.951600).

in current data there are variability between the two features V1 and V2 and it can be seen when comparing the minimum and maximum values of the two features(V1,V2) min(-7.042100, -13.773100),max(6.824800, 12.951600).

we have different features that are very different in scale and transform our data into a unified scale by **applying normalisation** before using the K-means algorithm algorithms (after which both our measurements would go from 0 to 1, preserving proportions, and allowing easier calculation between the two sets of features).

# Methods: How We Analyzed the Data

Data Preparation:

Data cleaning: We cleaned the data to ensure it was ready for analysis. in current data there are variability and in scale between the two features V1 and V2 and it can be seen when comparing the minimum and maximum values of the two features. We transformed our data into a unified scale by applying normalization after which both our measurements would go from 0 to 1, preserving proportions, and allowing easier calculation between the two sets of features.

2-Clustering: We used a method called K-means clustering to group the banknotes into two categories(clusters): genuine and fake.

3-Visualization: We created visual graphs to show how the banknotes were grouped into two clusters.

4-Stability Check: We ran the clustering method several times to make sure the results were consistent. (Figure 1)

Figure 1 :K-means clustering stability check
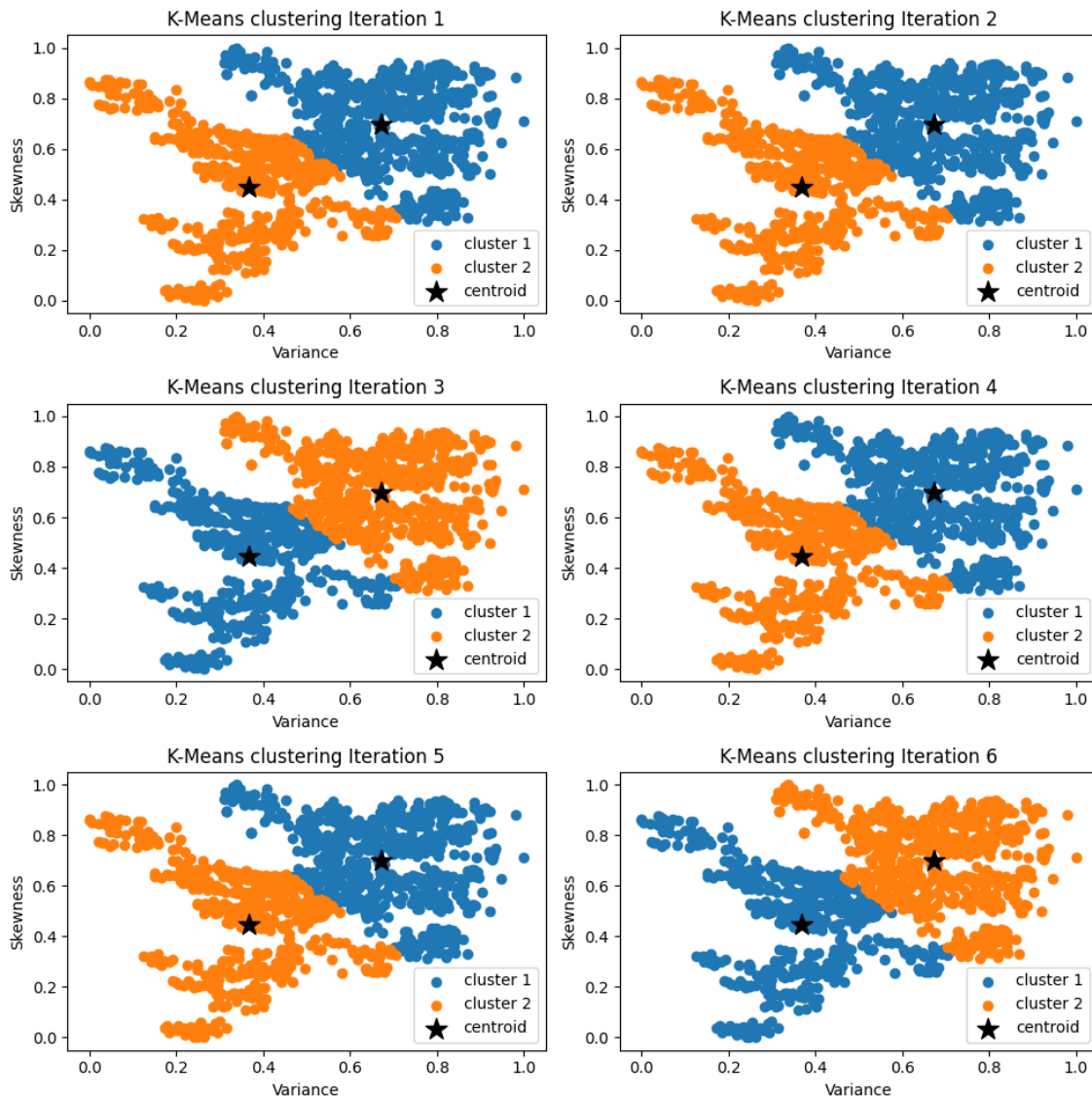
Summary of the Results

The K-means clustering method successfully grouped the banknotes into genuine and fake categories. The visual graphs showed a clear separation between the two groups, indicating that the features we used are effective for identifying fake banknotes. The results were consistent across multiple runs, showing that the method is reliable.

Recommendations

Based on our analysis, we recommend the following steps to automate the detection of forged banknotes:

1. **Implement Clustering**: Use the K-means clustering method to classify banknotes as genuine or fake. This method is effective and reliable.
2. **Regular Updates**: Continuously update the system with new data to ensure it remains accurate. Regular monitoring and retraining will help adapt to new counterfeit techniques.
3. **Integration**: Integrate the automated detection system with the bank's existing processes to streamline the identification and handling of fake banknotes.

Conclusion

Implementing an automated detection system using K-means clustering, along with other models, will significantly enhance the bank's ability to detect fake banknotes. This approach not only improves accuracy but also ensures the system remains adaptable to new counterfeit methods.