BANKNOTE AUTHENTIFICATION

DATASET DESCRIPTION
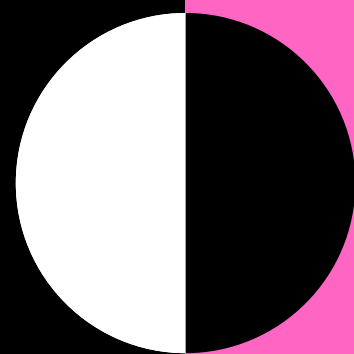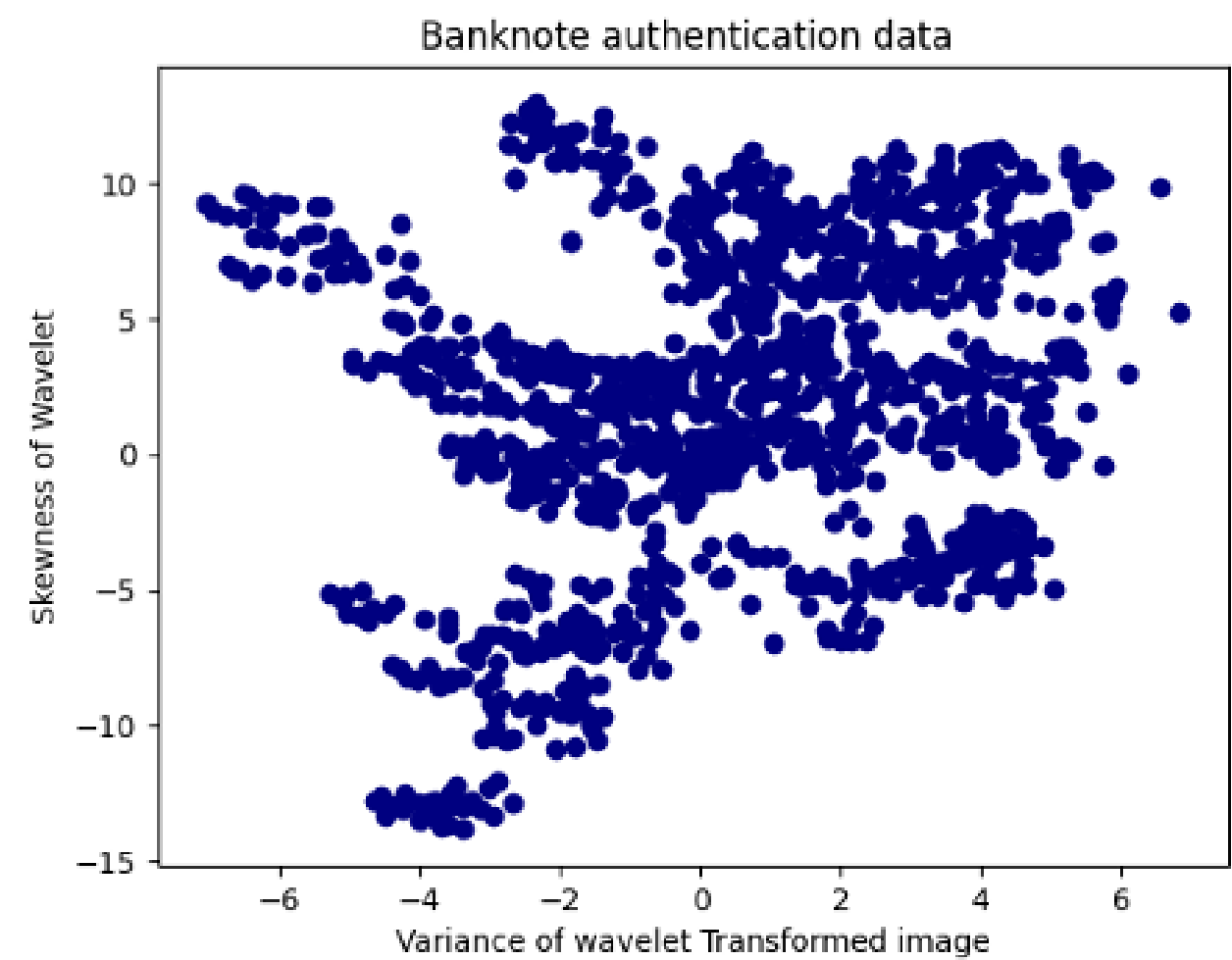

Banknote authentication data

In the first dataset, two variables are provided, V1(Variance of wavelet transformed image) and V2(Skewness of Wavelet Transformed image).

The data set consists of 1372 instances & no missing values.

Variable	Description	min	MAX	Mean	Median	Std
V1	Variance of wavelet transformed image	-7.0421	6.8248	0.433735	0.49618	2.841726
V2	Skewness of Wavelet Transformed image	-13.7731	12.9516	1.922353	2.31965	5.866907

EXPLORATORY DATA ANALYSIS

To display the variables, two histograms will be displayed:

Banknote authentication data
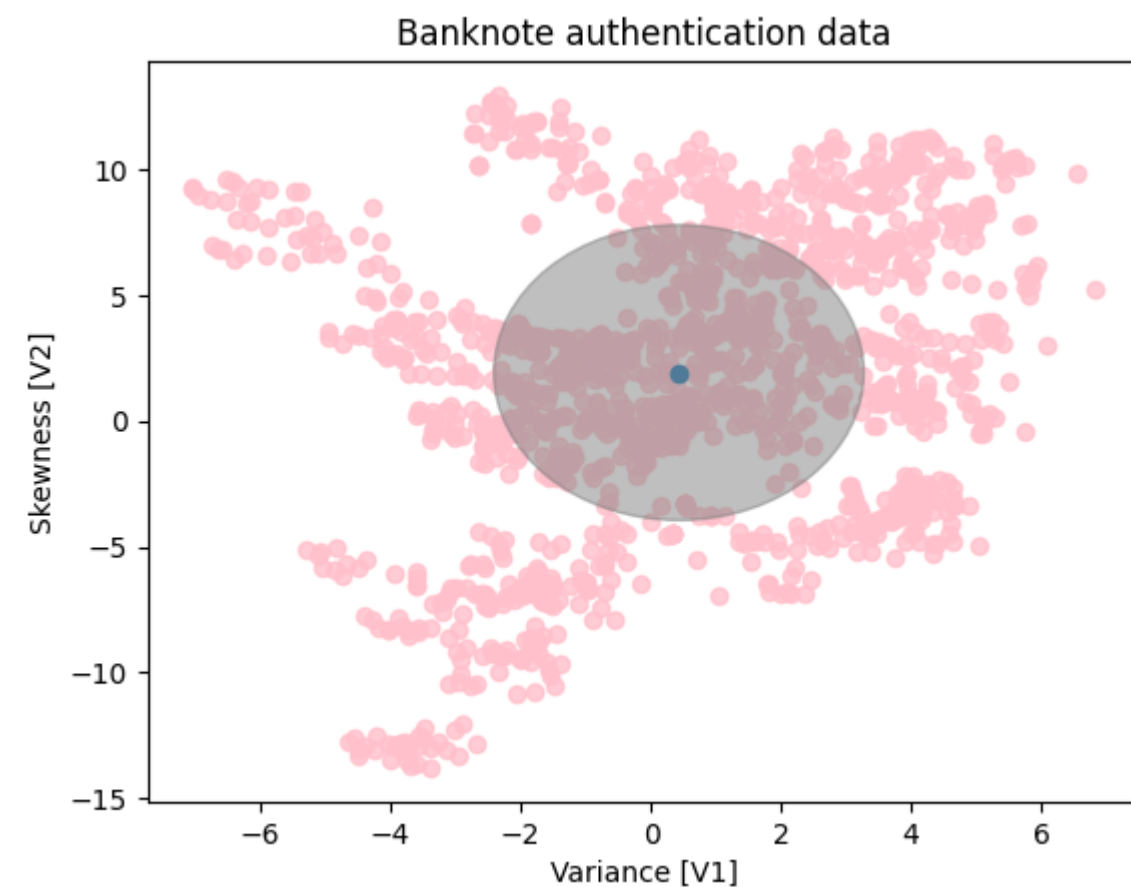

Banknote authentication data
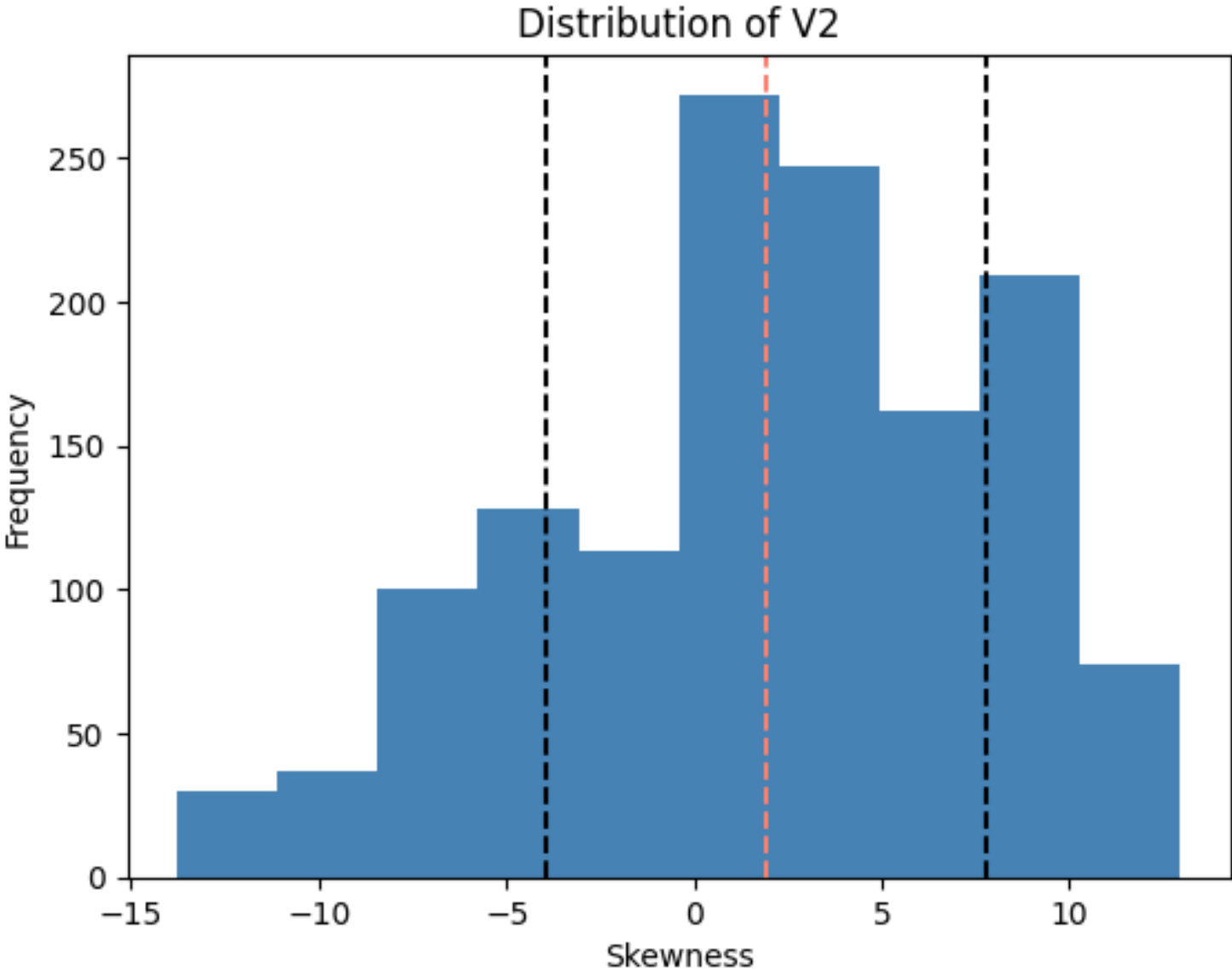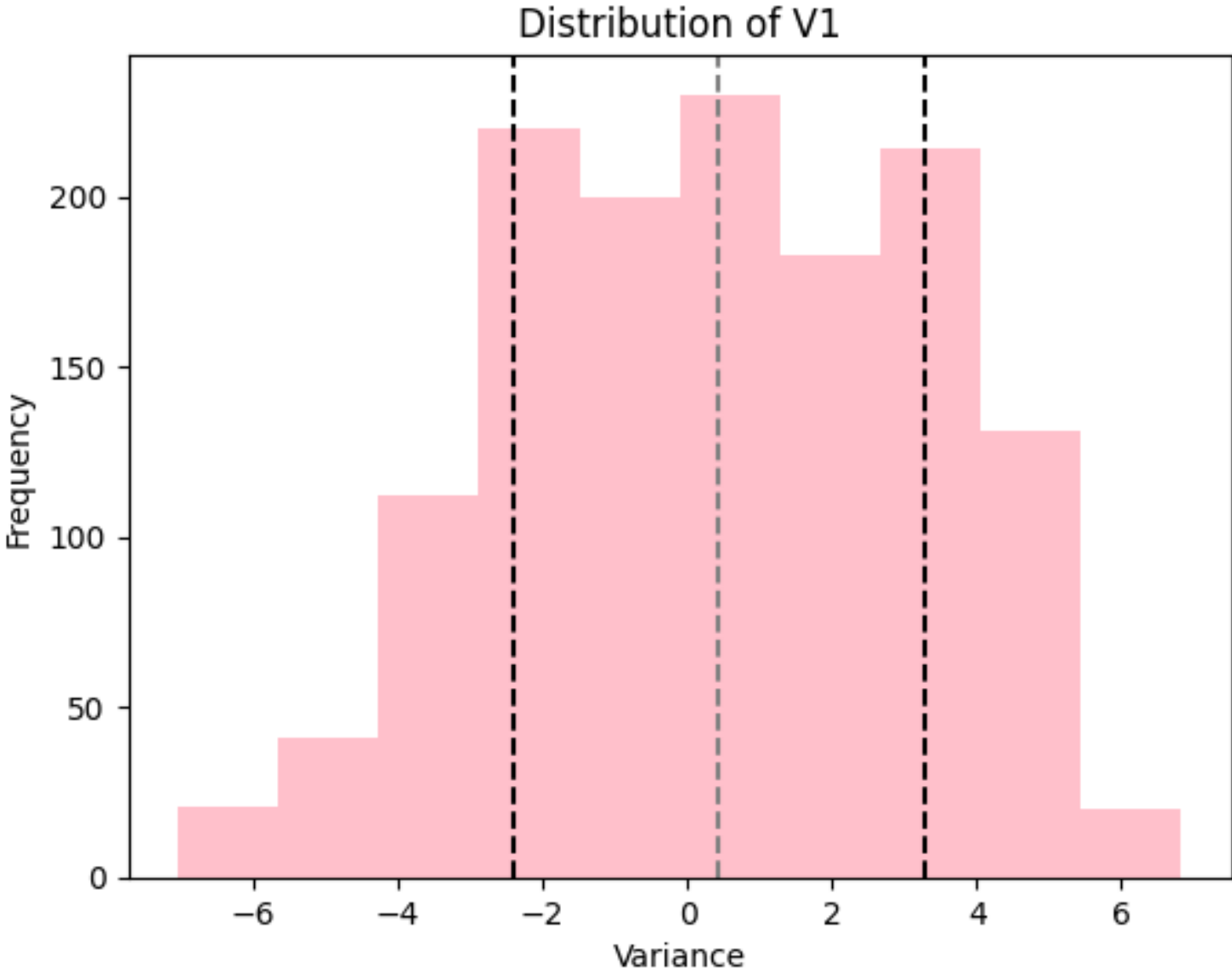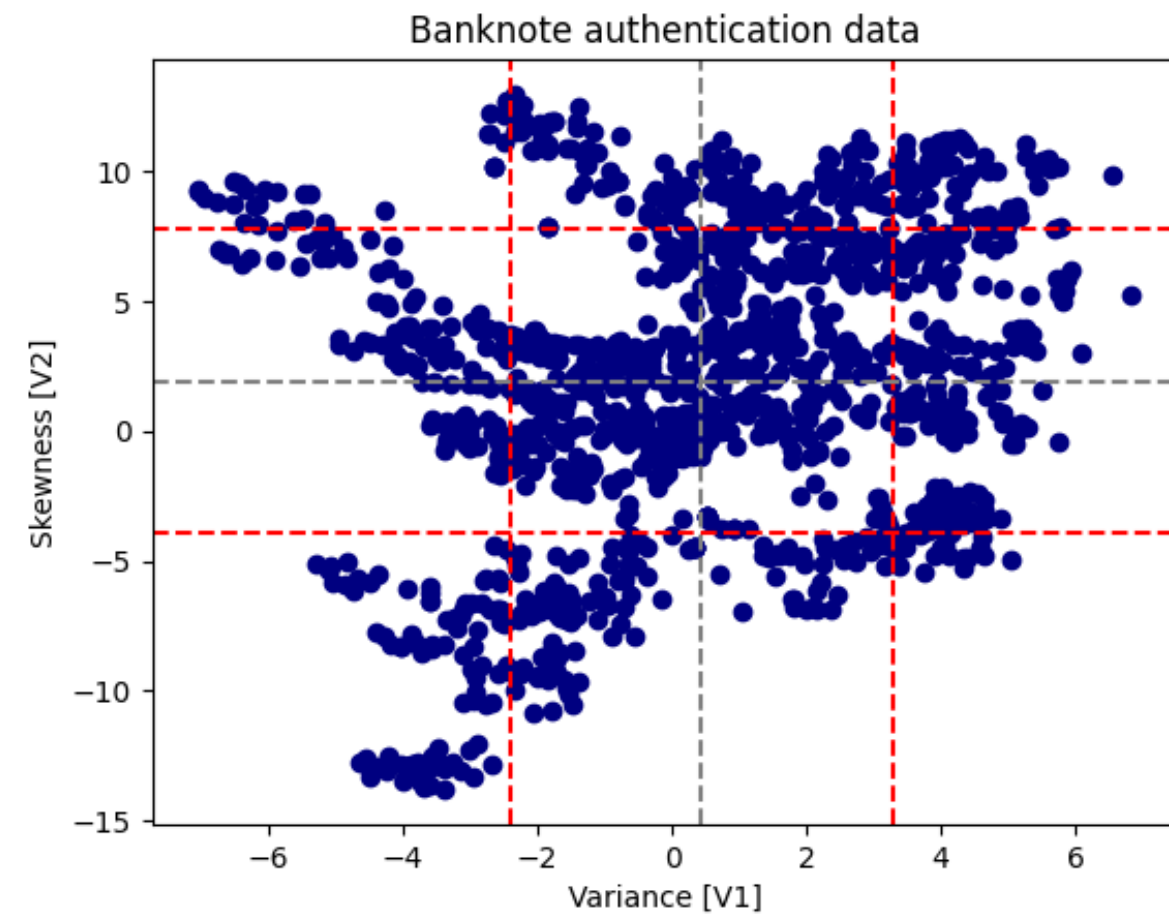
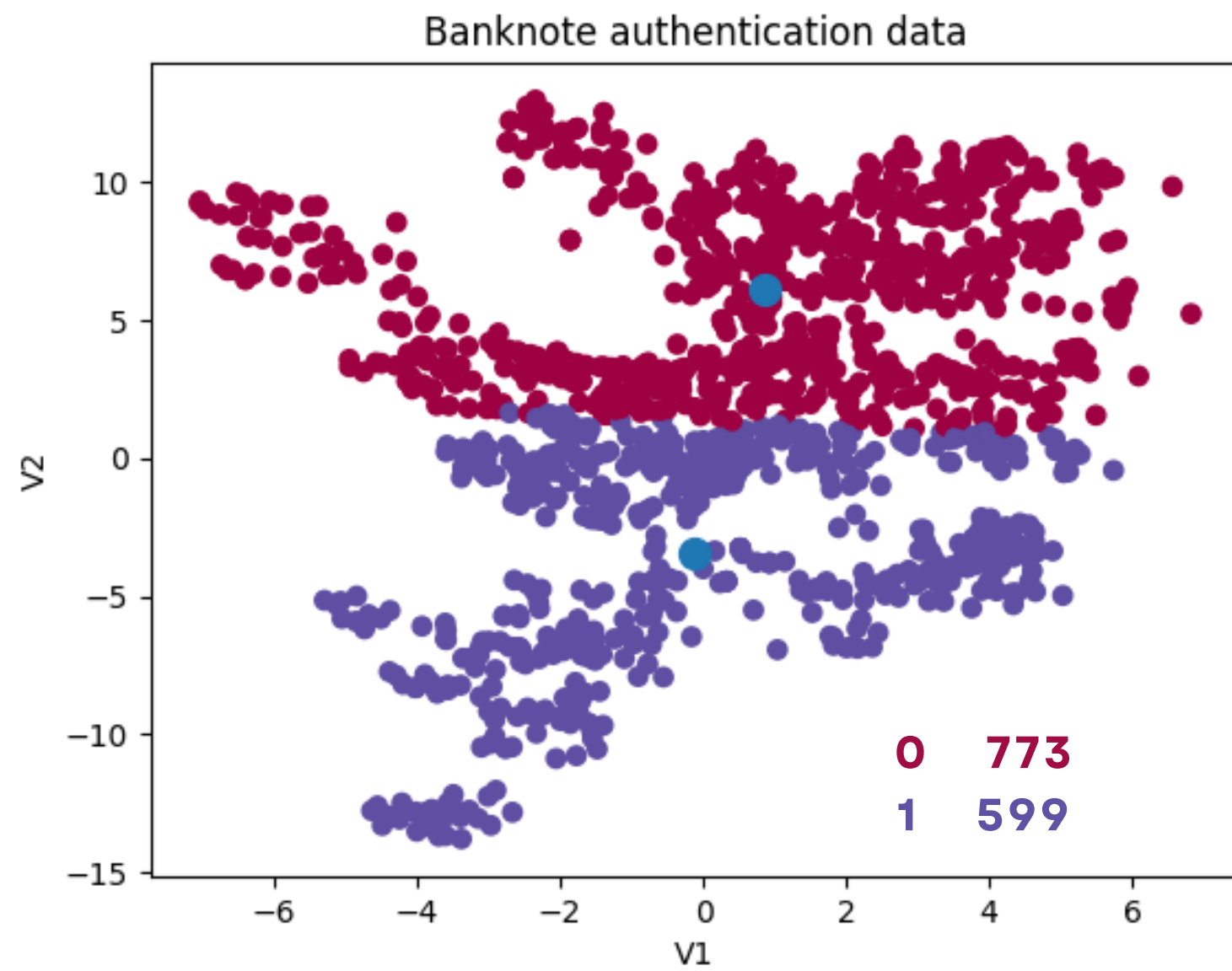## JUSTIFICATION OF THE SUITABILITY OF THE DATASET FOR K-MEANS CLUSTERING

K-means is suitable for large datasets, since this clustering algorithm is relatively fast vis-à-vis other algorithms.

Second, as all our variables are numeric, K-means is perfect to cluster all our data.

Third, data can be separated by variance and skewness.
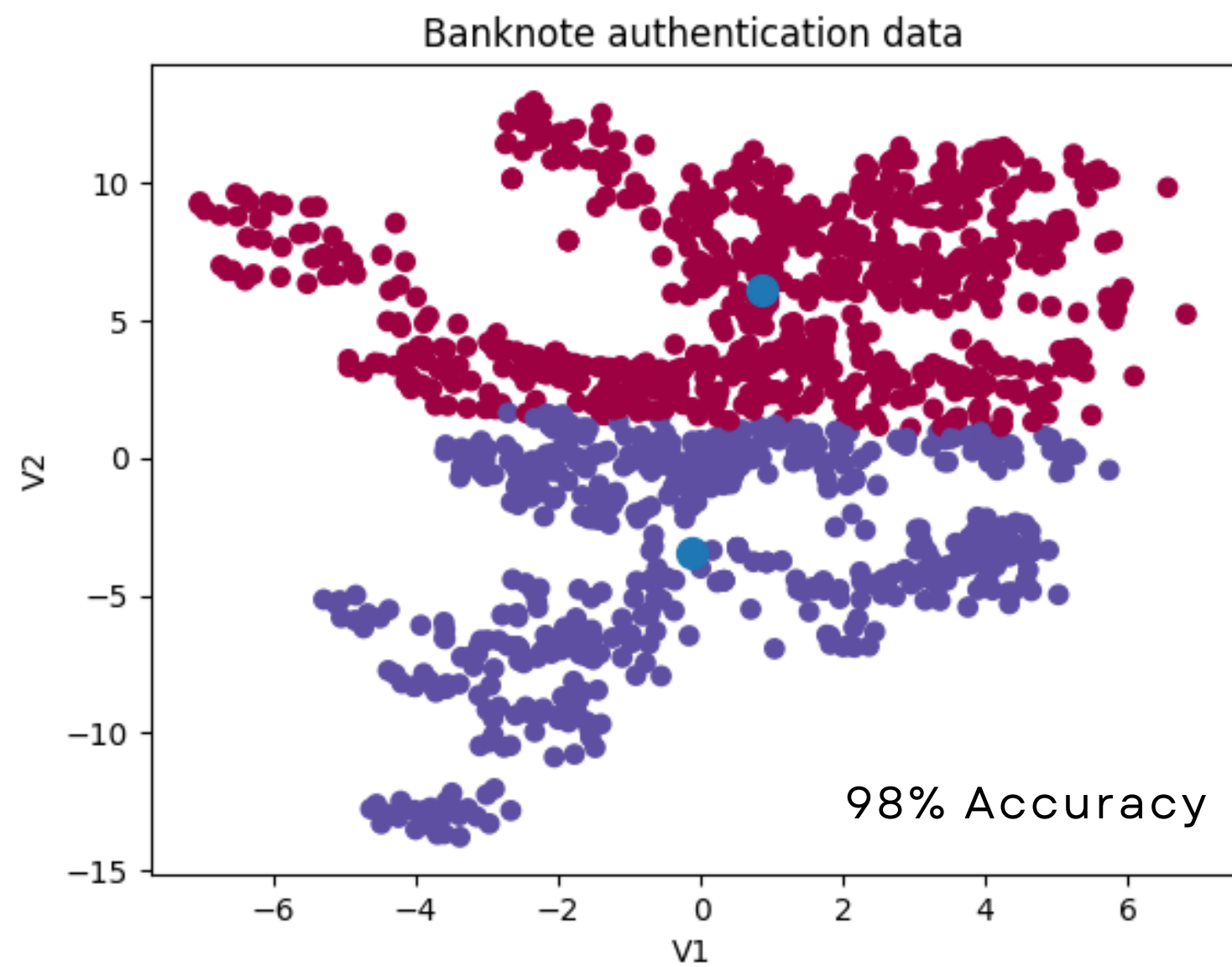
Although the algorithm assumes that clusters have a "spherical shape and similar variances", it will fit data beyond those borders.

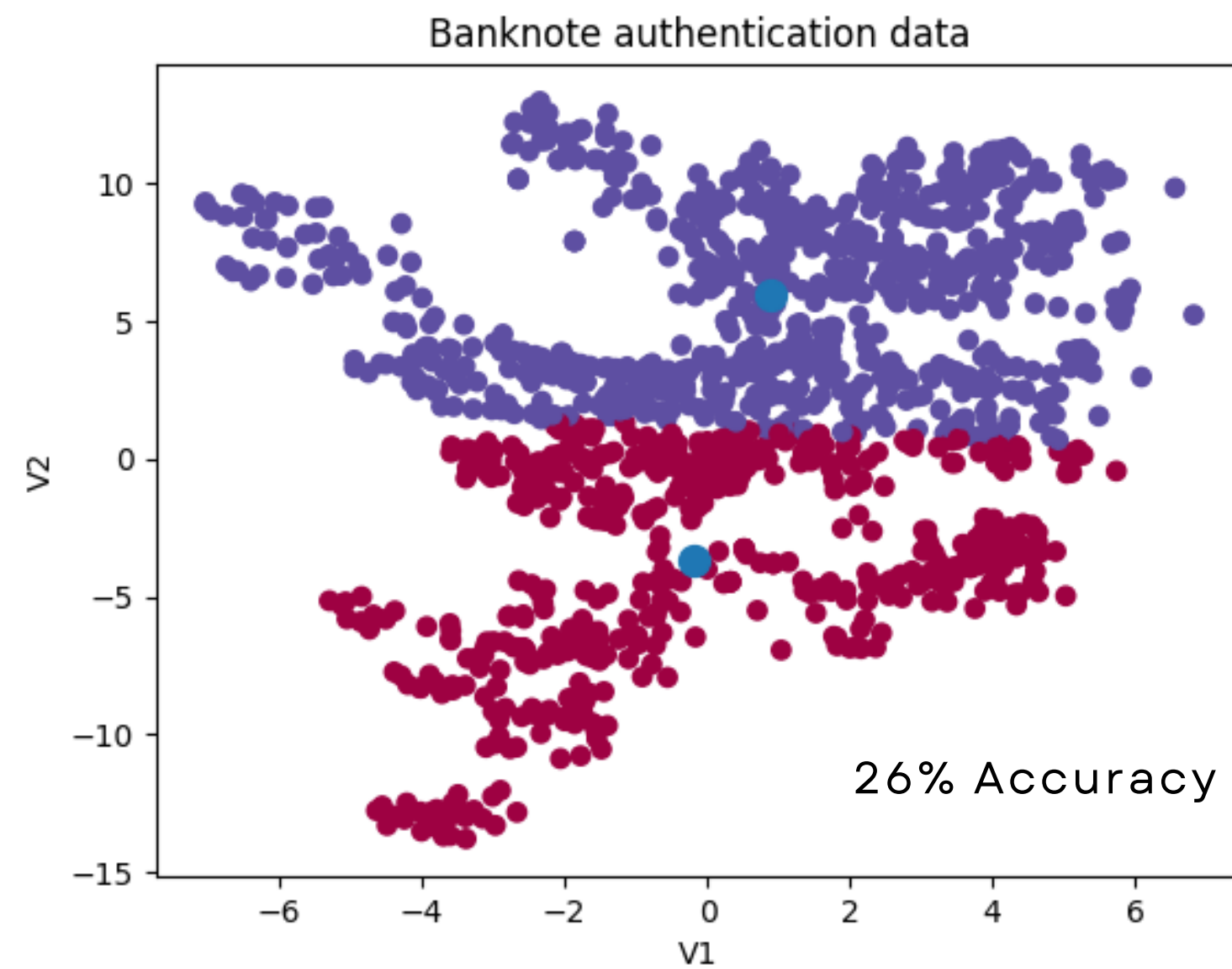Banknote authentication data

0   773
1   599

Prediction of 98% accuracy

Actual data:

0   761
1   610

Banknote authentication data

Actual data:

0	**761**
1	**610**

98% Accuracy

26% Accuracy

0	**773**	0	783	0	780	0	775	1	783	0	773	0	798	1	783	1	773
1	**599**	1	589	1	592	1	597	0	589	1	599	1	574	0	589	0	599

98.19	96.5	97.04	97.86	28.3	98.19	94	28.36	26.72

Clustering results

Banknote authentication data

Actual data:

0	**761**
1	**610**

98% Accuracy
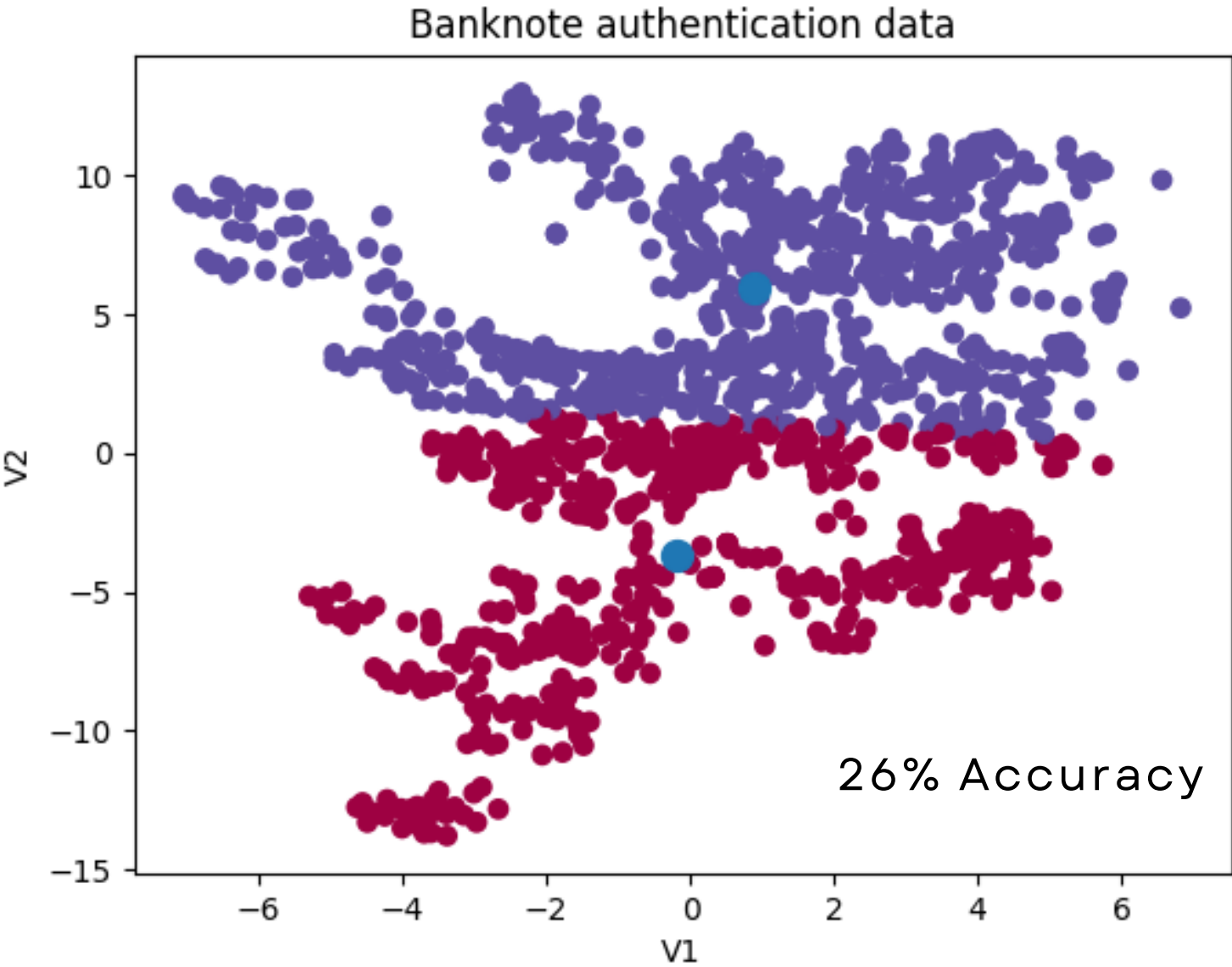
26% Accuracy

Trial	1	2	3	4	5	6	7	8	9
0	773	780	775	589	773	798	589	599	783
1	599	592	597	783	599	574	783	773	589
Total	1372	1372	1372	1372	1372	1372	1372	1372	1372
Accuracy (%)	98.1967213	97.0491803	97.8688525	28.3606557	98.1967213	94.0983607	28.3606557	26.7213115	96.557377