

Data Science Report for ABC Bank on forged banknotes

Introduction

ABC Bank (referred to as client) has asked us to use data science methods to do banknote forgery detection.

Problem

The client has been having bank notes forgery time and time again and needed a solution to detect watermark inside the banknote to determine whether it is real or not.

Dataset

The dataset is taken from OpenML website. The banknotes are scanned into images. Features are extracted using Wavelet Transform Tool. The dataset has two features: V1 (variance of Wavelet Transformed image) and V2 (skewness of Wavelet Transformed image).

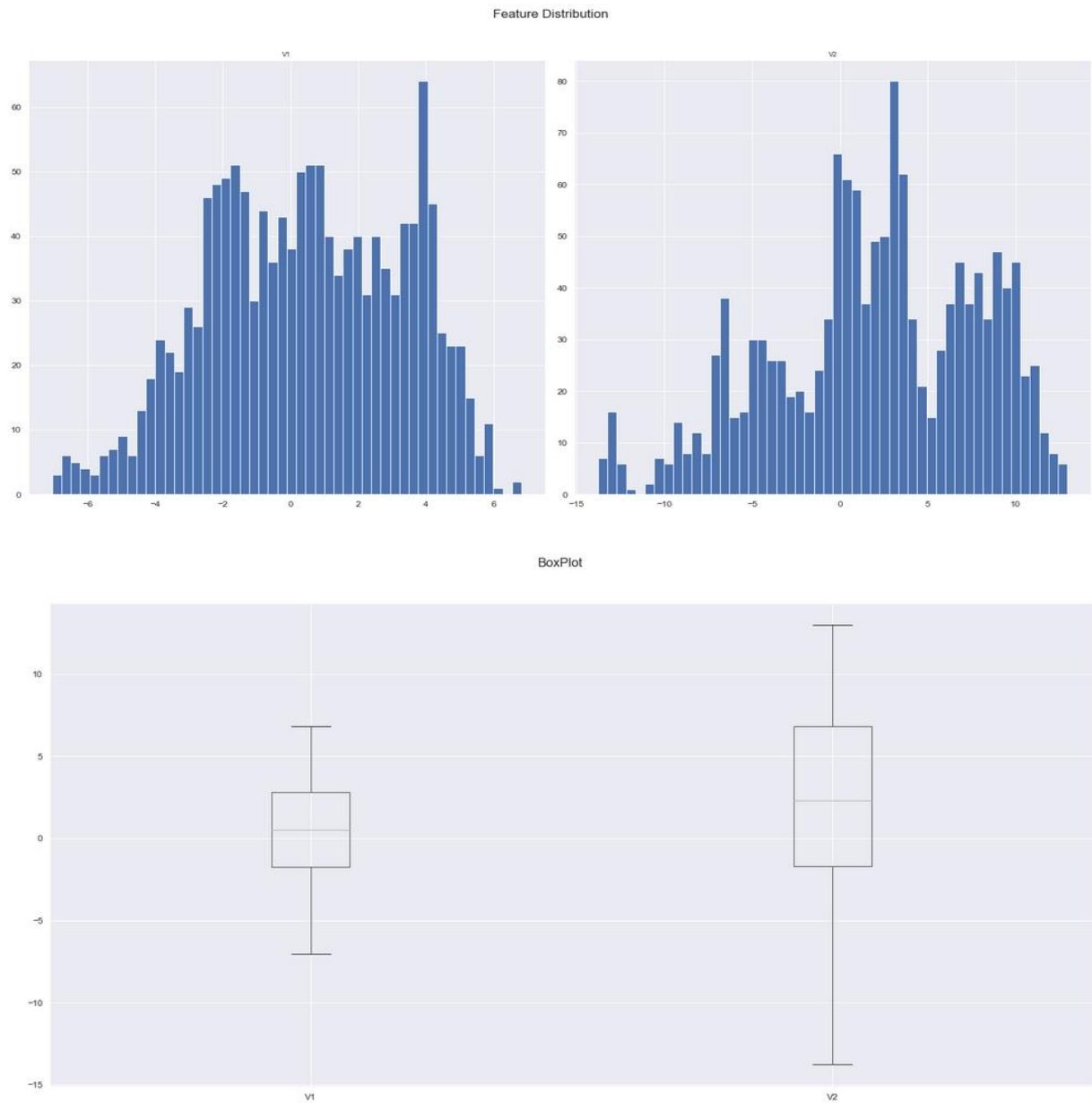
Data Analysis

The dataset contains 1372 observations and 2 features. Basic statistical analysis yields:

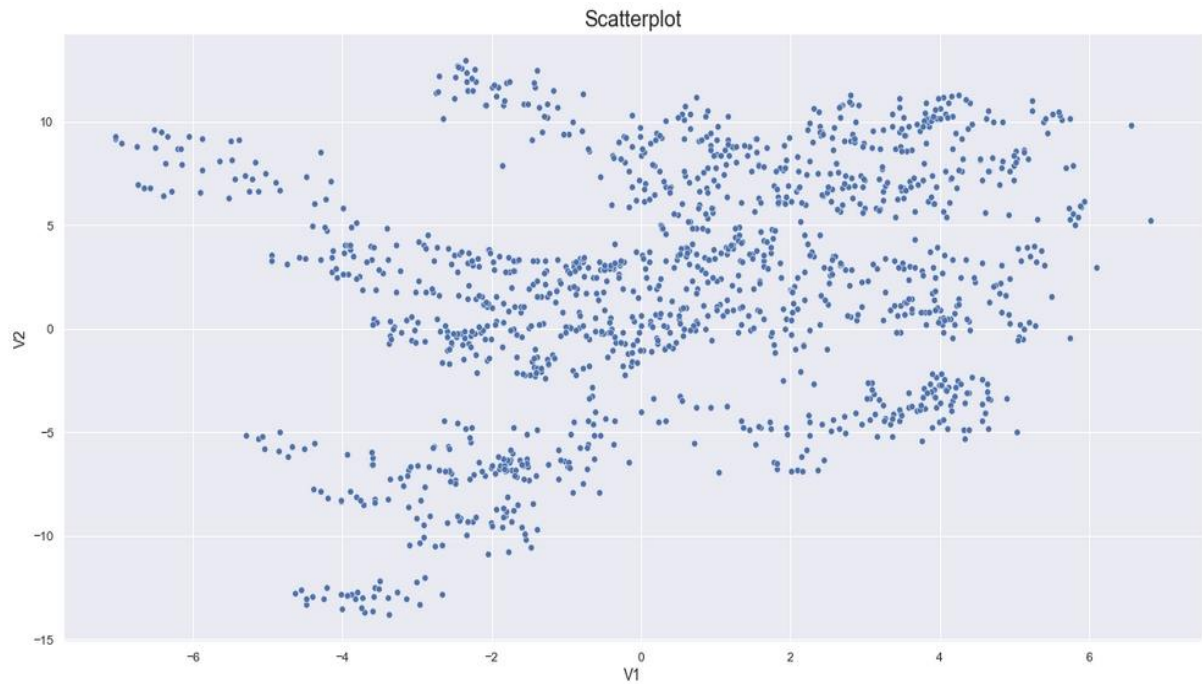
	V1	V2
count	1372.000000	1372.000000
mean	0.433735	1.922353
std	2.842763	5.869047
min	-7.042100	-13.773100
25%	-1.773000	-1.708200
50%	0.496180	2.319650
75%	2.821475	6.814625
max	6.824800	12.951600

Skewness has higher mean and standard deviation than variance measurements.

Both are not normally distributed as shown:



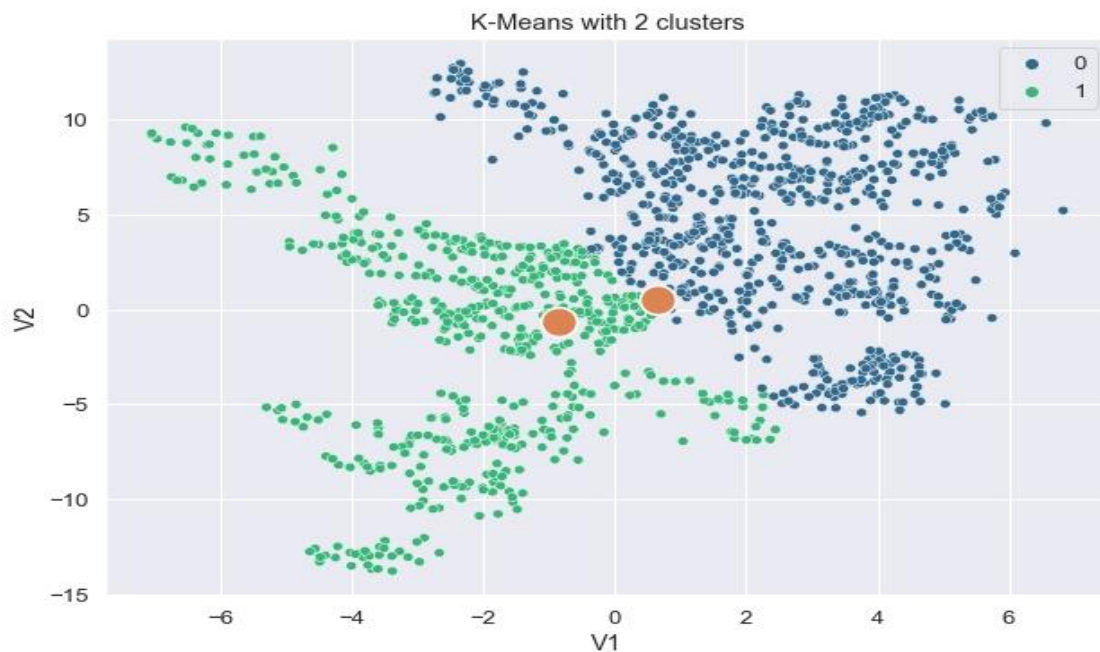
The scatterplot does not show any relationship between them:



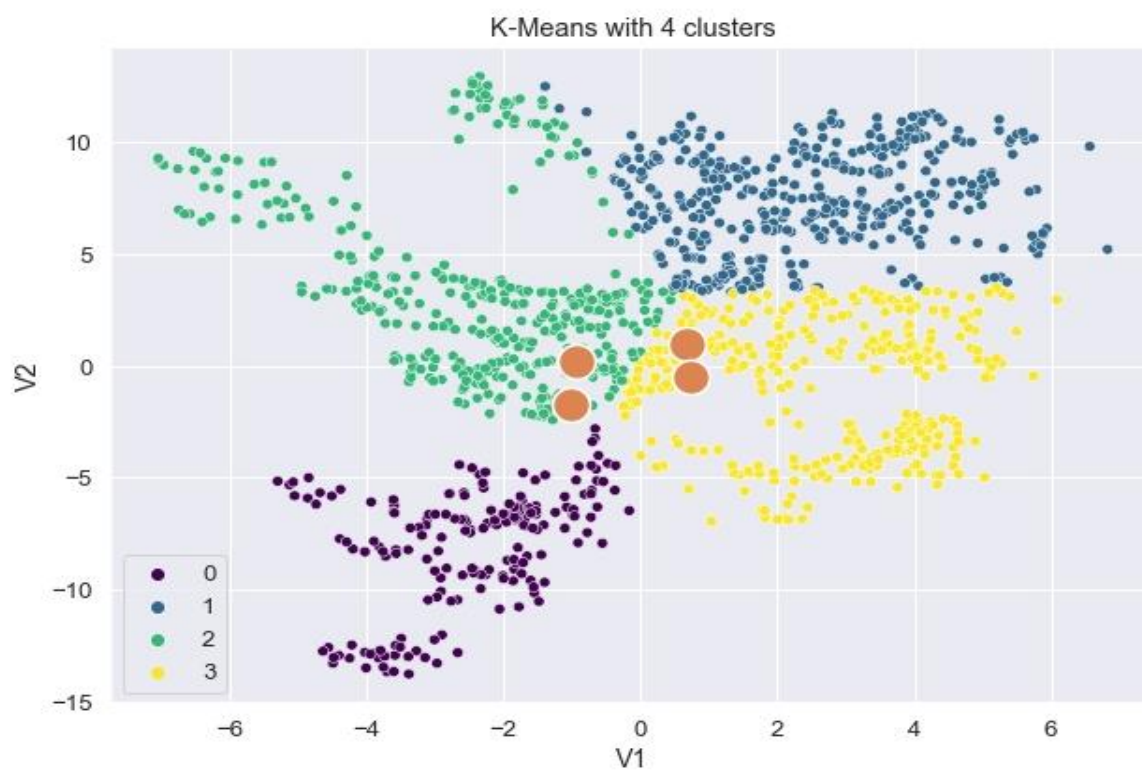
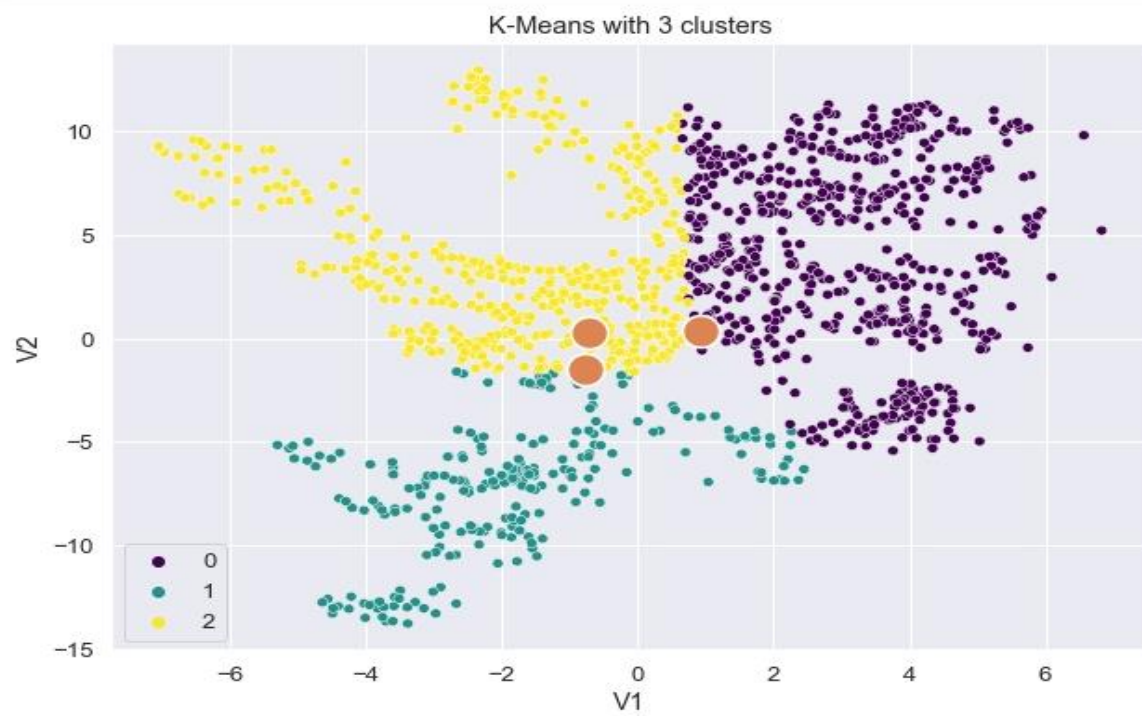
Modelling

Due to difficulty in finding patterns within the dataset, we used K-Means clustering method to find out.

We started with two clusters, there is a clear separation here.



We kept reiterating the same process by increasing each cluster sizes to see any effect.



Result

From the results, K-Means clustering is able to detect up to four different patterns within the dataset. Since our client is only keen on banknote is true or false, the problem is a binary type which is adequate.

Conclusion and Recommendation

We recommend extra features to be used like V3 and V4 and apply K-Means clustering to see further results.

We would like to try out using classification model if we are able to have samples of real and forged banknotes to be compared to this dataset to check for consistencies in data.