

PURPOSE

- To determine whether or not given banknotes are real or fraudulent, as to facilitate the decision of automating the detection of forged banknotes in your company.

DESCRIPTION

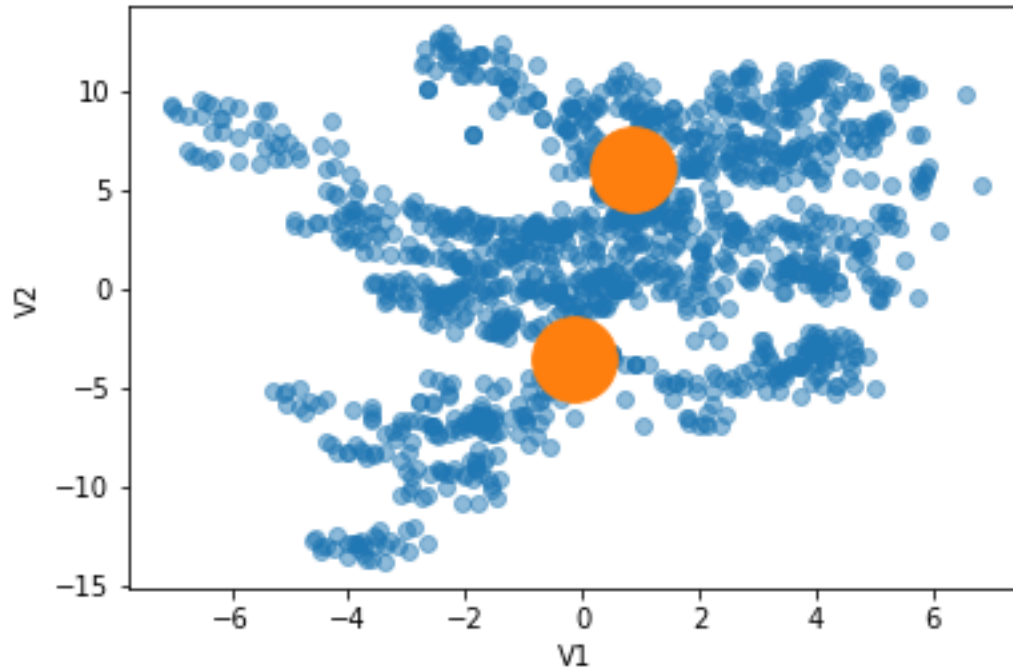
- Data from the OpenML community was used.
- The dataset contained multiple features, but only two of them were used in order to maintain simplicity: the variance and skewness.
- The dataset contained test results for multiple banknotes, making it reliable, and consisted of both real and fraudulent banknotes, giving equal priority to both.

METHODS USED

- The data analyzed includes sufficient data from both fraudulent and real banknotes.
- Images and features such as the two mentioned above was extracted for each note.
- The mean and standard deviations were calculated in order to determine just how much each note varied.
- A graph Variance vs Skewness was plotted to visualize the data, where each displayed point refers to a single bank note, and determine the effects of each feature in the genuineness of a banknote.
- A machine learning algorithm was used to determine the difference between a real and fraudulent banknote.
- The algorithm was run multiple times to make certain that the output is in fact reliable.

SUMMARY

- The analysis showed promising results as determined in the figure below.



- The algorithm used is trained to group data based on common patterns, in other words, it separates the real banknotes from the fraudulent ones, by grouping each separately.
- The two orange circles refer to each type of banknote, in other words one refers to the banknotes that're fraudulent and the other, the notes that are real, in the used dataset.
- The upper right circle refers to banknotes that're real.
- The lower left - the notes that're fraudulent.
- Based on this analysis, it can be said that the greater the variance and skewness, the more likely that the particular banknote is genuine.
- After running the analysis so many times, it was noticed that the above graph that displayed each time did not change, this can mean that the prediction is in fact correct. Why I say this is because the two orange circles kept remaining stable in their positions, so, based on the explanation above, this means that the two groups determined by the algorithm remained constant, or rather it was continuously confident that it had found the group of banknotes that're real and the group that're fraudulent.

RECOMMENDATIONS

- Based on the above discussion it's safe to say that it can be implemented as soon as possible.