

PROJECT DESCRIPTION

• Title of the project:

“Binary Classification for Banknote Authentication Dataset”

• Team Members:

- Tamanna Baig (tbaig@g.clemson.edu)
- Netra Inamdar (ninamda@g.clemson.edu)

• What data do you need and where will you get it?

The dataset we chose to work with is name Banknote Authentication Dataset. It represents the information regarding the authenticity of banknotes. This dataset was obtained from the UCI Machine Learning Repository and can be downloaded [here](#).

• Description of the problem you will be trying to solve? What question(s) are you trying to answer?

As mentioned above, the dataset represents the authenticity of banknotes. The dataset distinguishes between forged(1) and genuine(0) banknotes. Data were extracted from images that were taken from genuine and forged banknote-like specimens. For digitization, an industrial camera usually used for print inspection was used. The final images have 400x 400 pixels. Due to the object lens and distance to the investigated object gray-scale pictures with a resolution of about 660 dpi were gained. A Wavelet Transform tool was used to extract features from these images.

The dataset contains 5 features that can be separated as independent variables:

- X1: variance of Wavelet Transformed image (continuous)
- X2: skewness of Wavelet Transformed image (continuous)
- X3: curtosis of Wavelet Transformed image (continuous)
- X4: entropy of image (continuous)

And the outcome, dependent variable, Y : 0 for genuine, 1 for forged

The main questions we are trying to answer are: (but not limited to)

- To build a hypothesis function that can be used to predict the authenticity of the banknotes.
- To compare the results between the 2 methods and also with their existing counterparts to find out which algorithm will suit best.

• What machine learning algorithm(s) and models do you plan to use?

We plan to apply 2 of the project assignments covered in class.

- Binary Classification using kNN Algorithm
- Binary Classification using Logistic Regression

• How are you going to evaluate your results?

- Using Confusion Matrix the *Accuracy*, *Precision*, *Recall* and *F1 Score* of both the algorithms will be calculated.
- The results of the algorithms will be compared with the results of similar solution using existing ML libraries.