



# Binary Classification using Banknote Authentication Dataset

CPSC 6820: Special Topics: Hands-on Machine Learning

---

November 25, 2019

Clemson University

**Submitted by:** Tamanna Baig | Netra Inamdar

# Problem Description

---

- Banknote Authentication Dataset. This dataset was obtained from the UCI Machine Learning Repository and can be downloaded [here](#).
- A wavelet tool has been used to extract features from the images. Hence, the dataset has 5 distinct features that are explained in the Data Description.
- The aim of the project is to design a classification algorithm that will use the first 4 features as independent variables and predict a class outcome to distinguish between forged(0) and authentic(1) bank notes.

# Data Description

---

- The dataset is a text file with 1372 rows and 5 features.
- Data was extracted from images that were taken from genuine and forged banknote-like specimen.
- Each image was converted into grayscale with a resolution of about 660 DPI.
- 5 Features were extracted from these images using a Wavelet Transform tool.

## Attribute Information:

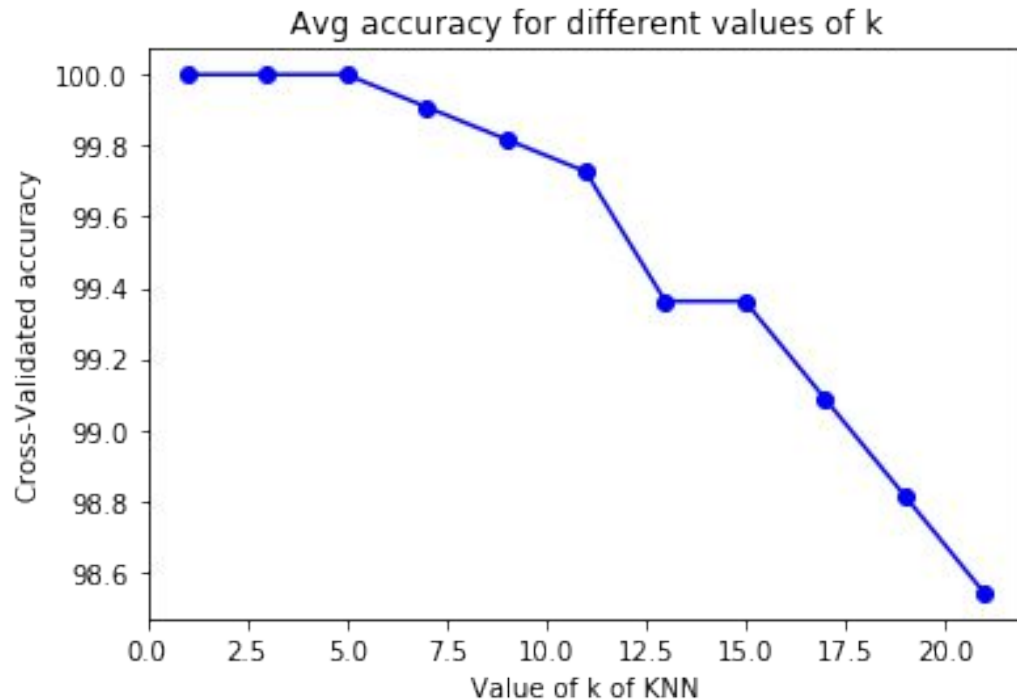
1. variance of Wavelet Transformed image (continuous)
2. skewness of Wavelet Transformed image (continuous)
3. curtosis of Wavelet Transformed image (continuous)
4. entropy of image (continuous)
5. class (integer)

Data Set Characteristics:	Multivariate	Number of Instances:	1372	Area:	Computer
Attribute Characteristics:	Real	Number of Attributes:	5	Date Donated	2013-04-16
Associated Tasks:	Classification	Missing Values?	N/A	Number of Web Hits:	204945

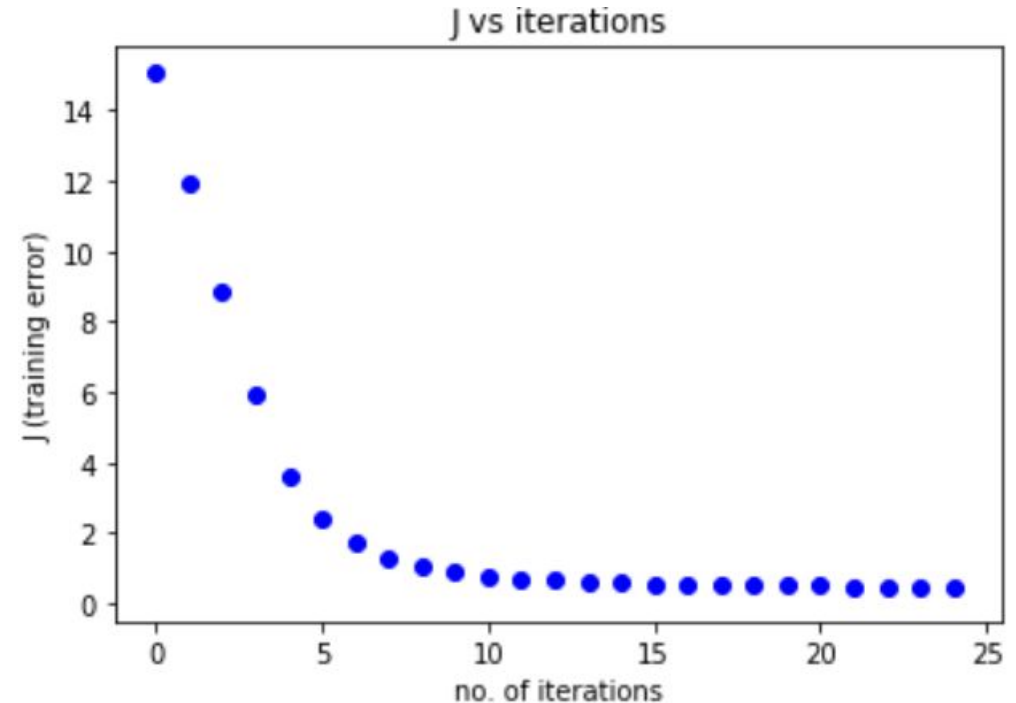
# Algorithms and Model

---

## Binary Classification using kNN algorithm:



## Binary Classification using Logistic Regression algorithm:



# Description for train, test and validation methods

---

- The dataset was split in a 80:20 ratio
  - **Train set:** 1097
  - **Test set:** 275
- **Feature Scaling:** Normalization/Scaling was applied to match the scales on all the features.
- **Cross Validation:** 5 fold cross validation was used in kNN algorithm along with train test split and normalization.

# Evaluation of Results

kNN		
Our Algorithm	Sklearn Algorithm	
Best k-value: 7	Best k-value: 16	For k-value = 7
TP, FP, TN, FN: 27 0 247 1	TP, FP, TN, FN: 149 0 126 0	TP, FP, TN, FN: 148 1 126 0
Accuracy: 0.996	Accuracy: 1.0	Accuracy: 0.9963
Recall: 1.0	Recall: 1.0	Recall: 1.0
Precision: 0.964	Precision: 1.0	Precision: 0.9921
F1 Score: 0.982	F1 Score: 1.0	F1 Score: 0.9960

Logistic Regression	
Our Algorithm	Sklearn Algorithm
TP, FP, TN, FN: 27 4 230 14	TP, FP, TN, FN: 148 1 125 1
Accuracy: 0.935	Accuracy: 0.9927
Recall: 0.659	Recall: 0.992
Precision: 0.871	Precision: 0.992
F1 Score: 0.753	F1 Score: 0.992