# Bank Note Authentication using SVM and Logistic Regression

Shrikant Madhukar Patil 202SP024, Tiwari Shubham Rajeshnath 202SP027
Department of Electronics and Communication Engineering
National Institute of Technology Karnataka, Surathkal
May 2021

## 1. Abstract:

There are many assets for a person or a country, banknotes are one of the most important assets of a country. Some miscreants introduce fake notes which bear a resemblance to original note to create discrepancies of the money in the financial market. It is difficult for humans to tell true and fake banknotes apart especially because they have a lot of similar features. Fake notes are created with precision, hence there is need for an efficient algorithm which accurately predicts whether a banknote is genuine or not. In this project we will use machine learning techniques to evaluate authentication of banknotes. We are comparing two algorithms here, State vector Machine and Logistic Regression and comparing there metrices we analyze both the algorithm.

Keywords: State Vector machine (SVM); logistic Regression; Confusion matrix; Recall, F1 score; Precision; Specificity; Accuracy

## 2. Introduction:

We often go to the bank to deposit some cash money, the cashier places banknotes in a machine which tells whether a banknote is real or not. This is a classification problem where we are given some input data and we have to classify the input into one of the several predefined categories. Rule-based as well as statistical techniques are commonly used for solving classification problems. Machine learning algorithms fall in the category of statistical techniques.

In this project, we explain the process of building a banknote authentication system using machine learning algorithms like state Vector Machine and Logistic regression. After that we will compare them to see which algorithm works better in case of Bank note authentication

### 2.1. Dataset

We have used dataset from UCI machine learning repository [1]. The dataset has five attributes out of which four attributes are real valued attributes and one attribute is target attribute. There are total 1372 instances. The target class contains two values namely 0 & 1, where 0 represents genuine banknotes and 1 represents fake banknotes. Data were extracted from images that were

taken from genuine and forged banknote-like specimens. Wavelet Transform tool were used to extract features from images.

## 2.2. Data

The description of the attributes is as follows.

TABLE 1 DATASET DESCRIPTION

| **Attribute name** | **Value Type** | **Description** |
|---|---|---|
| Variance | Double | It is a measure of the 'spread' of a distribution about its average value. |
| Skewness | Double | Skewness tells about the direction of variation of the lack of symmetry. |
| Curtosis | Double | Curtosis is a parameter that describes the peakedness of distribution |
| Entropy | Double | Image entropy is the amount of information which must be coded for, by a compression algorithm |
| Class | Nominal | Class contains two values namely 0 and 1 where 0 represents genuine banknotes and 1 represents fake banknotes |

We used Python libraries for the analysis of our dataset and trained the machine learning models from scratch. To import the dataset we used the Pandas library. For visualizing the dataset we used Seaborn library.

# 3. Implementation:

The experimental models have been setup using hold-out. Holdout method is one in which the data set is separated into two subsets (say 80:20 ratio) called the training set and the test set respectively. The training set is used to train the classifier while the testing set is used to estimate the error rate of the trained classifier.

## 3.1. Performance Measure

We have used the following metrics for measuring the performance of the algorithm.

### Confusion matrix:

It is used to evaluate the performance of the classifier. Each row in a confusion matrix represents an actual class, while each column represents a predicted class.

| | ***Predicted No*** | ***Predicted Yes*** |
|---|---|---|
| ***Actual No*** | **TN** | **FP** |
| ***Actual Yes*** | **FN** | **TP** |

Where,

TP= true positive (target class correctly classified)

TN= true negative (target class wrongly classified)

FP= false positive (non-target class correctly classified)

FN= false negative (non-target class wrongly classified)

1. **Precision:**
   Precision is the ability of a classifier not to label an instance positive that is actually negative [2]. It is given by

$$Precision = \frac{TP}{TP + FP}$$

2. **Sensitivity or recall:**
   This is the ratio of positive instances that are correctly detected by the classifier. It is also called True positive rate [2]. It is given as

$$Sensitivity = \frac{TP}{TP + FN}$$

3. **Specificity:**
   Specificity is the measure of how much ratio of non-target class is correctly classified by the particular classifier [2]. It is given as

$$Specificity = \frac{TN}{TN + FP}$$

4. **Accuracy :**
   Accuracy is the measure of how much ratio of whole dataset (including both target and non-target class) is correctly classified by the particular classifier [2]. It is given as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

5. **F1 score:**
   F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Intuitively it is not as easy to understand as accuracy, but F1 is usually more useful than accuracy, especially if you have an uneven class distribution.

## 3.2. Machine Learning algorithm

We have features which are extracted from images of real Notes with help of wavelet Transform .We have used State Vector Machine (SVM ) and Logistic Regression to

build Bank Note Authentication system that correctly classify the note as 0(real note) and as 1(fake note)

### 1. State Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification or regression problems. SVM takes all the data points in consideration and gives out a line that is called 'Hyperplane' which divides both the classes. SVM tries to make a decision boundary in such a way that the separation between the two classes is as wide as possible. If the data is non linearly separable then SVM makes use of kernel tricks to make it linearly separable [3].

y_hat = 0 if $w^T x+b < 0$ (0 means real note)

1 if $w^T x+b >= 0$ (1 means Fake note)

Objective function:

minimize (w,b) $\frac{1}{2}w^T w$

subject to $t^{(i)}(w^T x^{(i)}+b) >= 1$ fro i=1,2,……m

where t(i)= -1 for negative instances

= 1 for positive instances

### 2. Logistic Regression

Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes. It is one of the simplest ML algorithms that can be used for various classification problems such as spam detection, Diabetes prediction, cancer detection etc.Generally, logistic regression means binary logistic regression having binary target variables.

In logistic regression weighted sum of input is passed through the sigmoid activation function. The activation function takes any real values and converts them between 0 to 1. If the output given by a sigmoid function is more than 0.5, the output is classified as 1 & if is less than 0.5, the output is classified as 0 [3].

## 4. Results

The models where implemented with python from scratch to understand its working and also we used different classifier to understand which classifier works best for bank note authentication. We obtained the following result

### 1. Result of SVM

### Confusion Matrix-

Table 2 Confusion matrix of SVM

| | ***Predicted No*** | ***Predicted Yes*** |
|---|---|---|
| ***Actual No*** | 189 | 6 |
| ***Actual Yes*** | 0 | 148 |

Table 3 Metrics

| **Accuracy** | **Sensitivity** | **Precision** | **F1 score** | **Specificity** |
|---|---|---|---|---|
| 98.25% | 1 | 0.96 | 0.98 | 0.97 |

## 2. Result of Logistic Regression

**Confusion Matrix-**

Table 4 Confusion matrix of Logistic Regression

| | ***Predicted No*** | ***Predicted Yes*** |
|---|---|---|
| ***Actual No*** | 188 | 7 |
| ***Actual Yes*** | 0 | 148 |

Table 5 Metrics

| **Accuracy** | **Sensitivity** | **Precision** | **F1 score** | **Specificity** |
|---|---|---|---|---|
| 97.95 | 1 | 0.95 | 0.98 | 0.96 |

The above Table 2 and Table 3 summarizes the results of State Vector Machine, SVM classifier and Table 4 and Table 5 summarizes the results of logistic regression classifier. Comparing the table 3 and table 5 we see that SVM is giving accuracy of 98.25% and Logistic Regression is giving Accuracy of 97.95% and also Sensitivity of Both the classifier are same.

# 5. Conclusion:

In this project, we explained how we solved the problem of banknote authentication using machine learning techniques like SVM and Logistic Regression, which are built from scratch to understand it's working for basics. We compared both the algorithms in terms of performance we calculated Confusion matrix the comparing various metrics shown in the result table we concluded that the State vector Machine (SVM) algorithm is the best algorithm for banknote authentication with an accuracy of 98.25%. Therefore , rules given by State vector machine can be used to find out whether the given note is fake or not.

# 6. References

[1] https://www.kaggle.com/ritesaluja/bank-note-authentication-uci-data

[2] Chhotu Kumar, Anil Kumar dudyala, "Bank Note Authentication using Decision tree rules and machine learning techniques",2015 International Conference on Advances in Computer Engineering and Applications.

[3] Bishop, Christopher M., "Pattern Recognition and Machine Learning", Springer 2006.