# A

## Report

*On*

## Skill based Mini Project

*On*

# "Comparing the accuracy of various Classification Algorithms for Banknote Authentication(Data Mining)"

*For the course of*

## AI & ML (130617)

**Submitted By:**

**SHASHANK CHANDRAVANSHI (0901EE211104)**

**Under the guidance of**

## Dr. Nikhil Paliwal

**DEPARTMENT OF ELECTRICAL ENGINEERING**

माधव प्रौद्योगिकी एवं विज्ञान संस्थान, ग्वालियर (म.प्र.), भारत
**MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR (M.P.), INDIA**
Deemed to be University
(Declared under Distinct Category by Ministry of Education, Government of India)
NAAC ACCREDITED WITH A++ GRADE
**Race Course Road, Gola Ka Mandir, Gwalior, M.P. 474005**
Website: www.mitsgwalior.in

**25 APRIL 2024**

माधव प्रौद्योगिकी एवं विज्ञान संस्थान, ग्वालियर (म.प्र.), भारत
**MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR (M.P.), INDIA**
**Deemed to be University**
(Declared under Distinct Category by Ministry of Education, Government of India)
NAAC ACCREDITED WITH A++ GRADE

# CERTIFICATE

This is to certify that **SHASHANK CHANDRAVANSHI (0901EE211104)** studying in **MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE,** BATCH-2021-2025) have completed their Mini Skilled Based Project entitled **"Comparing the accuracy of various Classification Algorithms for Banknote Authentication(Data Mining)"** at **MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE** under my supervision.

It is further certified that they had attended required number of practical classes at **MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE,**GWALIOR for the completion of their Mini Skilled Based Project during 6th semester.

**DR. NIKHIL PALIWAL**

**Project Supervisor**

**माधव प्रौद्योगिकी एवं विज्ञान संस्थान, ग्वालियर (म.प्र.), भारत**
**MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR (M.P.), INDIA**
**Deemed to be University**
(Declared under Distinct Category by Ministry of Education, Government of India)
NAAC ACCREDITED WITH A++ GRADE

# CANDIDATE'S DECLARATION

We hereby declare that the work presented in this SKILLED BASED PROJECT entitled **"Comparing the accuracy of various Classification Algorithms for Banknote Authentication(Data Mining)"** which is being submitted in the partial fulfillment of the requirement for the award of degree of Bachelor of Engineering in Electrical Engineering is an authentic record of our own work carried out under the guidance of DR. NIKHIL PALIWAL, Electrical Engineering Department.

The matter presented in this project has not been submitted elsewhere by us for the award of any other degree/diploma.

**SHASHANK CHANDRAVANSHI**

**(0901EE211104)**

Date: 25/04/2024

Place: Gwalior

This is to certify that the above statement made by the candidates is correct to the best of my knowledge and belief.

Guided by

DR. NIKHIL PALIWAL
Department of Electrical Engineering
MITS, Gwalior

**माधव प्रौद्योगिकी एवं विज्ञान संस्थान, ग्वालियर (म.प्र.), भारत**
**MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR (M.P.), INDIA**
**Deemed to be University**
(Declared under Distinct Category by Ministry of Education, Government of India)
**NAAC ACCREDITED WITH A++ GRADE**

# ACKNOWLEDGEMENT

Engineers in all disciplines must acquire knowledge of project making. Student, in particular, will find 'project making' as an integral part of their studies that will infuse the spirit of doing practical work in them.

The satisfaction that accompanies the successful completion of any task would be incomplete without the mention of the people who made it possible whose constant guidance crowned our efforts with success.

We sincerely express our deep gratitude to the management of our college for giving us liberty to choose and to work on the most relevant project i.e. **"Comparing the accuracy of various Classification Algorithms for Banknote Authentication(Data Mining)".** We are thankful to DR. NIKHIL PALIWAL for ensuring that we have a smooth environment at the college and lab. At the very outset we would like to offer our never ending thanks to our project supervisor DR. NIKHIL PALIWAL who helped us with our project from the beginning till the end. His continuous surveillance over our work allowed us to work more efficiently.

SHASHANK CHANDRAVANSHI

(0901EE211104)

माधव प्रौद्योगिकी एवं विज्ञान संस्थान, ग्वालियर (म.प्र.), भारत
**MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR (M.P.), INDIA**
**Deemed to be University**
**(Declared under Distinct Category by Ministry of Education, Government of India)**
**NAAC ACCREDITED WITH A++ GRADE**

## ABSTRACT

There has been a significant increase in internet accessibility in the recent decade which has the potential to digitize the entire economy of the country. Despite this, physical currency in the form of bank notes remains the primary mode of transaction throughout the world. This has led to some unexpected day to day problems. Because of the massive advancement in printing technology, it has become very easy to produce counterfeit bank notes that look and feel similar to legitimate bank notes which makes it almost impossible to manually differentiate between the two. Thus our project aims at using the data extracted from the images of the bank notes to classify them as legitimate notes or counterfeit notes using certain classification algorithms such as Decision Tree, k-nearest neighbors (KNN), Random Forest and Support Vector Machine (SVM). Our project also compares and displays the accuracy of each algorithm used for the classification of the bank notes.

For our project we have used the bank note authentication data set. This data set contains data that were extracted from the images that were taken for the evaluation of an authentication procedure for banknotes. Using the wavelet transform tool, certain features were extracted from the images of bank notes. Our job is to apply various classification algorithms on this data set and predict whether the image taken is of legitimate bank note or of counterfeit bank note.

We have used four classification algorithms in our project namely Decision Tree, k-nearest neighbors (for k = 3 and k = 5), Random Forest and Support Vector Machine (SVM). After implementing the above algorithms and calculating their accuracy for our dataset, we found that the Decision Tree classifier had the highest accuracy (99.27%) whereas Support Vector Machine had the least accuracy (41.82%).

The main objective of undertaking this project is to implement some of the popular classification algorithms from scratch and identify the best algorithm that can be used to accurately differentiate counterfeit bank notes from legitimate notes.

माधव प्रौद्योगिकी एवं विज्ञान संस्थान, ग्वालियर (म.प्र.), भारत
**MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR (M.P.), INDIA**
**Deemed to be University**
(Declared under Distinct Category by Ministry of Education, Government of India)
NAAC ACCREDITED WITH A++ GRADE

## INDEX

माधव प्रौद्योगिकी एवं विज्ञान संस्थान, ग्वालियर (म.प्र.), भारत
**MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR (M.P.), INDIA**
**Deemed to be University**
(Declared under Distinct Category by Ministry of Education, Government of India)

# 1. INTRODUCTION

Despite the recent surge in the number of internet users in the country which is mainly attributed to the availability of cheap and affordable quality internet, the number of people utilizing the facility of online money transactions is trivial. Thus, exchange of physical currency in the form of bank notes remains the primary mode of transaction. With the advancements in printing technology it has become fairly simple to print counterfeit currency which looks and feels exactly similar to the legitimate ones. This has made it almost impossible to tell apart a counterfeit note from a real one manually just by looking at them. This has created a need for development of data mining models which can extract some important features from these bank notes and apply various classification algorithms on them to differentiate the legitimate notes from the counterfeit ones.

These models can be installed at various places where the inflow of currency is high, for example: banks, ATM's, vending machines, malls etc. The main societal challenge that our project is trying to overcome is to minimize the circulation of counterfeit notes in the market. The dataset that our project requires is the banknote authentication dataset which was created by taking the images of genuine and forged banknote like specimen and extracting certain features (like variance of Wavelet Transformed image, skewness of Wavelet Transformed image, curtosis of Wavelet Transformed image and entropy of image) from the images making use of various wavelet transform tools.

Finally, the aim of our project is to apply various supervised learning algorithms like decision tree, k-nearest neighbours, random forest and support vector machine on the given dataset to calculate and compare the accuracy of each of the following algorithms in classifying the bank notes as genuine or forged in order to find the best classification algorithm that can be used in our model to solve our problem.
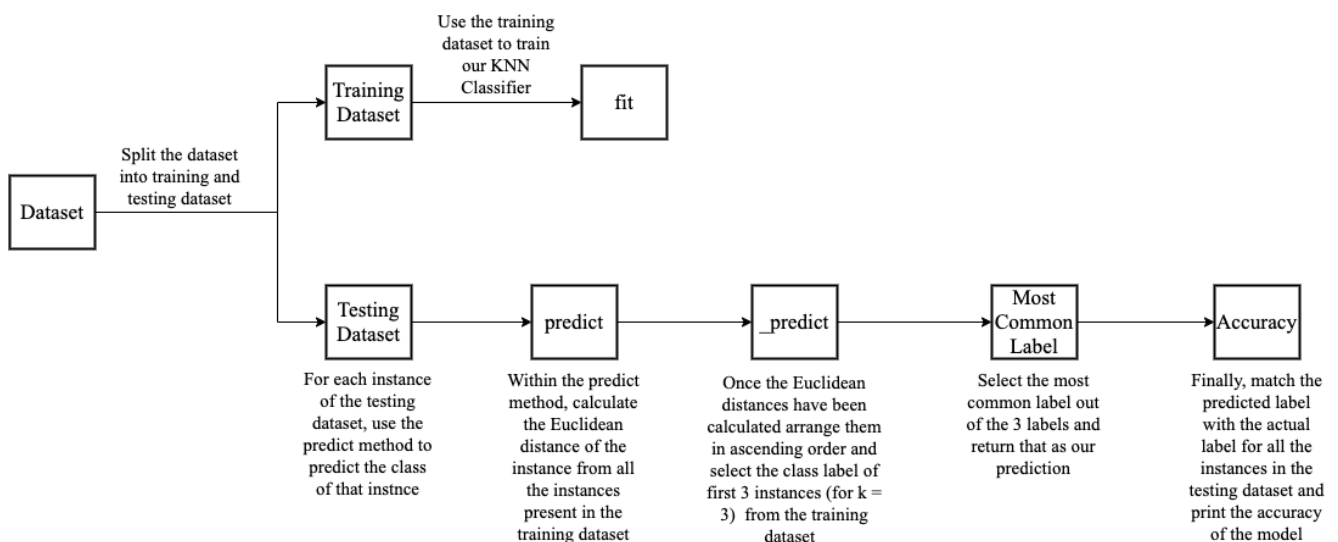
# 2. DATASET

The dataset that we have used in our project is the banknote authentication data set taken from the UCI machine learning repository. This dataset was created by capturing the images of genuine as well as forged banknote-like specimens using an industry grade print inspection camera. The features (like variance of Wavelet Transformed image, skewness of Wavelet Transformed image, curtosis of Wavelet Transformed image and entropy of image) were extracted from these photos by using various Wavelet Transform Tools. The dataset consists of 5 attributes in total where the first four attributes are the extracted features namely: variance of Wavelet Transformed image, skewness of Wavelet Transformed image, curtosis of Wavelet Transformed image and entropy of image and the fifth attribute is the target class which has two class labels 0 (represents genuine notes) and 1 (represents forged notes). The dataset has 1372 instances in total out of which instances 1 to 762 belong to class 0 and instances 763 to 1372 belong to class 1. Thus, the legit:counterfeit ratio in the dataset is 56:44.

माधव प्रौद्योगिकी एवं विज्ञान संस्थान, ग्वालियर (म.प्र.), भारत
**MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR (M.P.), INDIA**
**Deemed to be University**
(Declared under Distinct Category by Ministry of Education, Government of India)

## Table 1. Dataset Description

| Attribute Name | Value Type | Description |
|---|---|---|
| Variance of Wavelet Transformed Image | Continuous | Variance is the measure of how a pixel varies from its neighbouring pixels |
| Skewness of Wavelet Transformed Image | Continuous | Skewness measures how asymmetrical the image is |
| Curtosis of Wavelet Transformed Image | Continuous | Kurtosis is a measure of whether the data is heavy tailed or light-tailed relative to a normal distribution |
| Entropy of Image | Continuous | The entropy or average information of an image is a measure of the degree of randomness in the image |
| Class | Integer | Class 0 represents genuine notes whereas class 1 represents counterfeit notes |

## 3. ARCHITECTURE

### K - Nearest Neighbours Classifier:



- Split the dataset into training and testing dataset.
- Use the training dataset to train our KNN Classifier.
- For each instance of the testing dataset, use the predict method to predict the class of that instance.
- Within the predict method, calculate the Euclidean distance of the instance from all the instances present in the training dataset.

**माधव प्रौद्योगिकी एवं विज्ञान संस्थान, ग्वालियर (म.प्र.), भारत**
**MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR (M.P.), INDIA**
**Deemed to be University**
**(Declared under Distinct Category by Ministry of Education, Government of India)**
NAAC ACCREDITED WITH A++ GRADE

- Once the Euclidean distances have been calculated, arrange them in ascending order and select the class label of first 3 instances (for k = 3) from the training dataset.
- Select the most common label out of the 3 labels and return that as our prediction.
- Finally, match the predicted label with the actual label for all the instances in the testing dataset and print the accuracy of the model

**Random Forest Classifier:**
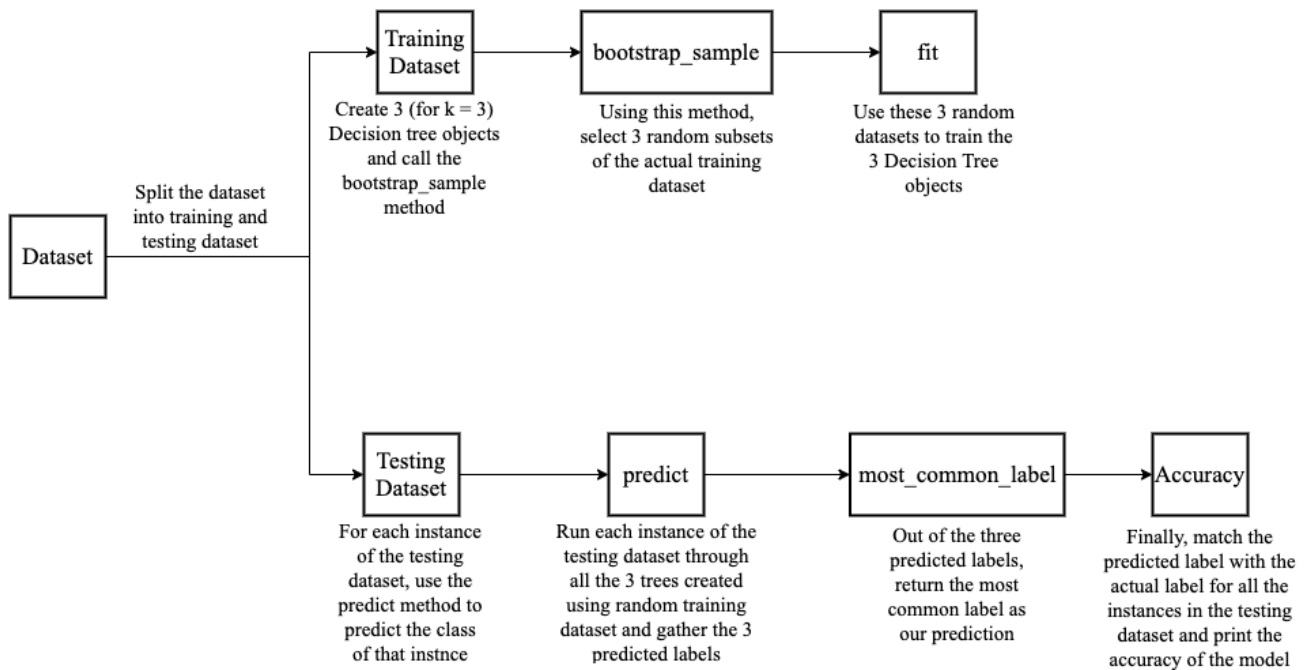


- Split the dataset into training and testing dataset.
- Create 3 (for k = 3) Decision tree objects and call the bootstrap_sample method.
- Using this method, select 3 random subsets of the actual training dataset.
- Use these 3 random datasets to train the 3 Decision Tree objects.
- For each instance of the testing dataset, use the predict method to predict the class of that instance.
- Run each instance of the testing dataset through all the 3 trees created using the random training dataset and gather the 3 predicted labels.
- Out of the three predicted labels, return the most common label as our prediction.
- Finally, match the predicted label with the actual label for all the instances in the testing dataset and print the accuracy of the model.

माधव प्रौद्योगिकी एवं विज्ञान संस्थान, ग्वालियर (म.प्र.), भारत
**MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR (M.P.), INDIA**
**Deemed to be University**
(Declared under Distinct Category by Ministry of Education, Government of India)
**NAAC ACCREDITED WITH A++ GRADE**

**Input and Output of learning models:**

1. Decision Tree Classifier:

- Input: Banknote Authentication Dataset, maximum depth of the tree
- Output: Accuracy (in %) of the model in differentiating fake notes from legitimate notes.

2. K - nearest neighbours Classifier:

- Input: Banknote Authentication Dataset, number of neighbours to consider (k value) while predicting the class label
- Output: Accuracy (in %) of the model in differentiating fake notes from legitimate notes.

3. Random Forest Classifier:

- Input: Banknote Authentication Dataset, number of trees to train
- Output: Accuracy (in %) of the model in differentiating fake notes from legitimate notes.

4. Support Vector Machine:

- Input: Banknote Authentication Dataset
- Output: Accuracy (in %) of the model in differentiating fake notes from legitimate notes.

## 4. EXPERIMENT

- In our project, we have implemented four different classification algorithms (Decision Tree, KNN, Random Forest, Support Vector Machine) from scratch.
- The Decision Tree algorithm has been implemented with the following parameters: maximum depth of the tree is 10, information gain is calculated using Entropy of the nodes and minimum number of samples required to split the tree is 2.
- The k-nearest neighbour algorithm has been implemented with the following parameters: number of nearest neighbours considered while predicting the class is 3 and 5 (both individually), the distance between two points is measured using the euclidean distance.
- The Random Forest algorithm has been implemented with the following parameters: number of decision trees to be constructed is 3, maximum depth of each tree is 10, information gain is calculated using Entropy of the nodes, minimum number of samples required to split the tree is 2, taking same instance of training dataset more than once to create a random subset training dataset is allowed.
- The Support Vector Machine algorithm has been implemented with the following parameters: the learning rate is 0.001, the lambda parameter is 0.01 and the number of iterations are 1000.
- Finally, the accuracy of all these algorithms are calculated and displayed in a tabular fashion for easy comparison.

**माधव प्रौद्योगिकी एवं विज्ञान संस्थान, ग्वालियर (म.प्र.), भारत**
**MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR (M.P.), INDIA**
**Deemed to be University**
**(Declared under Distinct Category by Ministry of Education, Government of India)**
**NAAC ACCREDITED WITH A++ GRADE**

## 5. RESULT

- Upon implementation of 4 different classification algorithms from scratch and comparing their accuracies in classifying a banknote as legitimate or counterfeit, we achieved the following result:

- The most accurate classification algorithm for the given dataset was the Decision Tree Classifier with an accuracy of 99.27%.

- The second best algorithm for the given dataset was the Random Forest Classifier with an accuracy of 98.91%.

- The third best algorithm for the given dataset was the K-Nearest Neighbours Classifier with 'k' set to 3 having an accuracy of 96.0%.

- The fourth best algorithm for the given dataset was the K-Nearest Neighbours Classifier with 'k' set to 5 having an accuracy of 95.64%.

- And the worst performing algorithm for the given dataset was the Support Vector Machine having an accuracy of only 41.82%.

### Table 2. Algorithms and their Accuracies

| Sr. No. | Algorithm | Accuracy (%) |
|---------|-----------|--------------|
| 1. | Decision Tree Classifier | 99.27 |
| 2. | Random Forest Classifier | 98.91 |
| 3. | KNN Classifier (k = 3) | 96.0 |
| 4. | KNN Classifier (k = 5) | 95.64 |
| 5. | Support Vector Machine Classifier | 41.82 |

## 6. CONCLUSION

In our project, we have implemented four different classification algorithms from scratch namely: Decision Tree Classifier, K-Nearest Neighbours Classifier, Random Forest Classifier and Support Vector Machine Classifier and calculated the accuracies of each algorithm on the banknote authentication dataset.

After analyzing the accuracies of these algorithms in differentiating the counterfeit notes from the legitimate ones, we conclude that that Decision Tree Classifier is the best algorithm with an accuracy of 99.27% and Support Vector Machine is the worst algorithm having an accuracy of 41.82% for the given dataset.

Thus, Wavelet Transform Tool along with Decision Tree Classifier can be used to create a model which can accurately predict a given note as legitimate or counterfeit. Hence, this model can be installed at various places where the inflow of currency is high, for example: banks, ATM's, vending

माधव प्रौद्योगिकी एवं विज्ञान संस्थान, ग्वालियर (म.प्र.), भारत
**MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE, GWALIOR (M.P.), INDIA**
**Deemed to be University**
(Declared under Distinct Category by Ministry of Education, Government of India)
**NAAC ACCREDITED WITH A++ GRADE**

machines, malls etc. which could help in minimizing the circulation of counterfeit notes in the market to a great extent.

In future, this work can be extended by implementing some of the other classification algorithms apart from the ones already implemented in this project or by varying some of the parameters for the above implemented algorithms so as to achieve higher accuracy.