**Project Title :** This model attempts to make an accurate identification about a currency note, whether it is authentic or inauthentic using four specifications of the photograph of the currency note.

**Problem Definition :**

The project deals with the classification of the notes of a particular currency (whether authentic or inauthentic) using some features of the note. Here, the algorithms like logistic regression, support vector classifier, decision trees, gaussian naïve Bayes and self- implemented k- means algorithms are used and after comparison amongst them any one can be selected for further work in other aspects.

**Design and implementation of the project :**

**Using R :**

1. The dataset is first loaded into a data frame of nine variables, the first four variables basically contains all the data, the fifth contains the pre- predicted column and the fifth and the sixth are for the computation purposes.
2. The first four variables are normalised to lie between 0 and 1.
3. Then an overlay plot is made using the first variable as x- axis and the others features as y- axes independently with the x.
4. A self developed k-means algorithm is applied and the confusion matrix is created and the accuracy comes around 42.45%, which is same when the inbuilt k-means function is applied.
5. The cost function is then calculated for a maximum limit of 200 clusters and its corresponding graph is then plotted.
6. Using this model, if we give the required inputs (four of the specifications) of the photograph of a currency note, we can identify it as authentic or inauthentic.

This model can be made better using other algorithms which are implemented in python, but here the main display is about the self implemented k-means computation and its equivalent result and the plotting of the cost function for the various number of clusters (k) taken.

**Using Python:**

The preprocessing of both the train and test dataset is performed in prior. All the feature columns are check and filled up with appropriate values in case if any missing values are found for both the train and test dataset in the preprocessing step. The train data is then trained with various models such as Logistic Regression, Support Vector Classifier, Decision Tree Classifier and Gaussian Naïve Bayes Classifier. The best model is then selected on the basis of accuracy score and used with test data to predict if the input currency note is authentic or not.

The given problem is implemented as shown in the flowchart given below:

**Import all the required libraries**

**Load the datasets**

**The train dataset is then fitted by using various classification models**

**Select the best model based on the accuracy score**

**Use the test data on the selected model and predict if the input currency note is authentic or not**

**Conclusion :** This model will can be used as a basic model to build a sophisticated model for the detection of duplicate notes in notes checking machines used in banks, etc. Using some more features and observation this model can be made finer and finalised to be used as a full- time working model. Thus is it is a very dynamically extendible model, but as for the scope of this project it has been concluded by testing its efficiency with the aforementioned machine learning algorithms and can be surely refined to the professional level.