



Analysis and Comparison of Data Mining Tools and Techniques for Classification of Banknote Authentication

A. K. Shrivastava
Department of IT
Dr. C. V. Raman University
Bilaspur, India

Priyanka Gupta
Department of IT
Dr. C. V. Raman University
Bilaspur, India

Abstract: Banknote authentication is very important and challenging task for every banks and financial sectors to secure the information from unauthorized person. Banknote authentication avoids fraud or misuses of financial system and protects the financial data from unauthorized users. Classification play very major role to classify the authentic and counterfeit notes. In this paper we have presented the classification accuracy of different classifiers and compared the accuracy of classifiers using three data mining tools. In first case compared the accuracy of classifiers using Rapid miner and WEKA (Waikato Environment for Knowledge Analysis) data mining tools where Random Forest gives better accuracy as 99.30% in case of WEKA data mining tool. In second case compared the accuracy of classifiers using Rapid miner and Tanagra data mining tool where K-NN gives 99.56% of accuracy in Tanagra data mining tool. Finally compared the classification accuracy using WEKA and Tanagra data mining tool where MLP gives 99.1% of accuracy in case of WEKA data mining tool.

Keywords: Banknote Authentication, Classification, WEKA, Rapid Miner, Tanagra.

I. INTRODUCTION

Banknote authentication (Nife N. I. et al., 2016) is an important challenge for the various private and commercial banks in order to keep the strength of the financial system around the world. The numbers of fake notes are increasing due to advancement of digital image processing. This authentication problem is not only facing by the banks but also facing by most of the peoples. The main serious problem is how to identify and categorize the genuine and counterfeit notes. There are various researchers and authors have used data mining, soft computing, statistical techniques, various feature selection and optimization techniques to develop the robust and efficient classifiers for classification of authenticate and non authenticate notes. There are various authors have worked in the field of bank note authentication. C. R. Jyothi et al. [3] have suggested back propagation neural network to identify the authentic notes. N. I. Nife et al. [1] used various data mining techniques like Decision trees-j48, Multi Layer Perceptron, EM algorithm, K Mean Algorithm and Naïve Bayes on banknote authentication. They found that Multi Layer Perceptron algorithm is superior to other techniques. R. Ghosh et al. [7] have suggested a technique of currency recognition using Neural Network. B. S. Prasanthi [4] have used image processing techniques for authentication of Indian paper currency. They have used six features including identification mark, security thread, watermark, numeral, floral design, micro lettering from the image of the currency. S. Sen Sarma [6] used Genetic Algorithm and Neural based approach for Banknote authentication. The proposed model is compared with well-known MLP-FFN classifier. The proposed NN-GA model is superior to other models under current study. S. Jadhav et al. [2] using

advance image processing approach to extract the color, aspect ratio and the unique identification for identification of authentic notes. E. Kaya et al.[9] have suggested Artificial Neural Network to perform clustering process and classify the banknote authentication. The proposed method is capable to identify authentic note with high accuracy. R. Mirza et al. [10] have suggested image processing techniques to identify the authentication of Indian paper currency using identification mark, security thread and watermark. The feature is extracted using proposed edge based segmentation by Sobel operator and computation time is less in whole process. S. E. Ali et al. [8] have discussed various methods and tools to identify the authentic notes. They have also discussed objectives and significant of methods and tools for banknote authentication.

II. PROPOSED ARCHITECTURE

The architecture of proposed research work as shown in Figure 1. The main objective of this research work is to identify and classify the authentic and non authentic notes. In this research work, we have collected banknote authentication data set from UCI repository and analysis of classification techniques with three data mining tools like WEKA, Tanagra and Rapid Miner with 10-fold cross data partition. We have used different classification techniques and compared the classification accuracy of classifiers with two different data mining tools as shown in figure 1. Finally select the best classifier with suitable data mining tool.

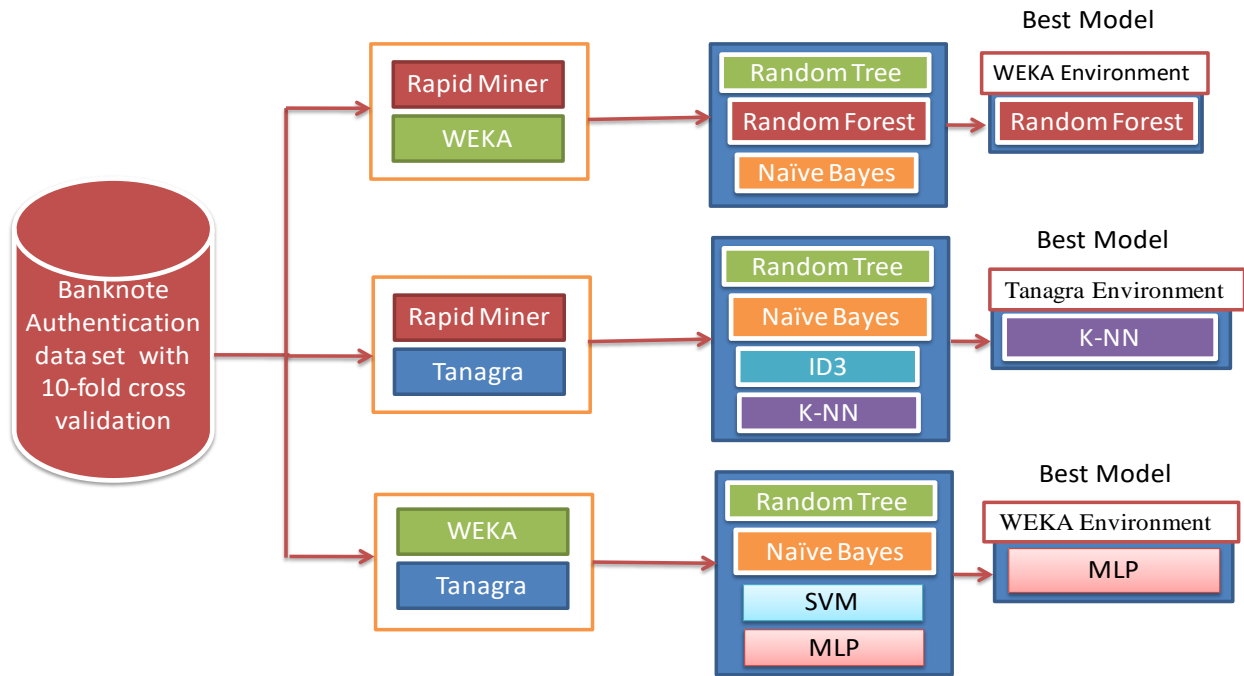


Figure1. Architecture of proposed model

A. Data Set

We have collected banknote authentication data set from UCI repository [14]. The data set consist 1372 instances, 4 features and 1 classes having authenticated and non-authenticated notes. The data set no contains missing value.

B. Data Mining Tools

In this research work, we have used WEKA [15], Tanagra [16] and Rapid Miner [17] data mining tools for analysis and classification of data. These tools are used for regression, classification, clustering, association rule mining, and attribute selection. It provides extensive support for the whole process of experimental data mining, including preparing the input data, evaluating learning schemes statistically, and visualizing the input data and the result of learning [12].

C. Data Mining based Classification Technique

Data mining is extracting of knowledge from large amount of data. Data mining based classification techniques play very important role to classify the data. Classification process consists into two steps: training and testing. Training data is used to train the classifier and testing data is used to test the trained classifier.

• Decision Tree

Decision tree [13] is probably the most popular data mining technique. The most common data mining task for a decision tree is classification. The principle idea of a decision tree is to split our data recursively into subsets so that each subset contains more or less homogeneous states of our target variable (predictable attribute). At each split in the tree, all input attributes are evaluated for their impact on the predictable attribute. When this recursive process is completed, a decision tree is formed. In this research work we have used Random Tree, Random Forest and Iterative

Dichotomizer 3 (ID 3) classification techniques for classification of banknote authentication.

• Naïve Bayes

Naïve Bayes [5] is a statistical classifier and it is also called simple Naïve Bayes classifier. Bayesian classifiers are statistical classifiers. Naïve Bayes classifier to be comparable in performance with decision tree and selected neural network classifiers. Bayes classifiers have also exhibited high accuracy and speed when applied to large databases.

• Support Vector Machine (SVM)

Support Vector Machines [5] is a promising new method for the classification of both linear and nonlinear data. It uses a nonlinear mapping to transform the original training data into a higher dimension. Within this new dimension, it searches for the linear optimal separating hyperplane. With an appropriate nonlinear mapping to a sufficiently high dimension, data from two classes can always be separated by a hyperplane. Although the training time of even the fastest SVMs can be extremely slow, they are highly accurate, owing to their ability to model complex nonlinear decision boundaries. They are much less prone to overfitting than other methods. The support vectors found also provide a compact description of the learned model.

• K-Nearest Neighbor (K-NN)

The k-nearest-neighbor [5] method is labor intensive when given large training sets. It has since been widely used in the area of pattern recognition. Nearest-neighbor classifiers are based on learning by analogy, that is, by comparing a given test tuple with training tuples that are similar to it. The training tuples are described by n attributes. Each tuple represents a point in an n -dimensional space. In this way, all of the training tuples are stored in an n -dimensional pattern space. When given an unknown tuple, a k-nearest-neighbor

classifier searches the pattern space for the k training tuples that are closest to the unknown tuple. These k training tuples are the k “nearest neighbors” of the unknown tuple. “Closeness” is defined in terms of a distance metric, such as Euclidean distance.

- **Multilayer Perceptron (MLP)**

MLP [11] is a development from the simple perceptron in which extra hidden layers (layers additional to the input and output layers, not connected externally) are added. More than one hidden layer can be used. The network topology is constrained to be feed forward, i.e., loop-free. Generally, connections are allowed from the input layer to the first (and possibly only) hidden layer, from the first hidden layer to the second and so on, until the last hidden layer to the output layer. The presence of these layers allows an ANN to approximate a variety of non-linear functions. The actual construction of network, as well as the determination of the number of hidden layers and determination of the overall number of units, is sometimes of a trial-and-error process, determined by the nature of the problem at hand. The transfer function generally a sigmoid function.

III. RESULT AND DISCUSSION

This research work is carried out using three data mining tools i.e. WEKA, Tanagra and Rapid miner for analyzing the banknote authentication. This experiment is divided into three sections: firstly compare the accuracy of classifiers using Rapid miner and WEKA data mining tools. Secondly, compare the accuracy of classifiers using Rapid miner and Tanagra and finally, compare the accuracy of classifiers using WEKA and Tanagra data mining tools.

In first section, we have compared the accuracy of Random Tree, Random Forest and Naive Bayes algorithm using Rapid miner and WEKA data mining tools. In case of both data mining tools, Random Forest gives better classification accuracy. The Random Forest gives 57.54% and 99.3% of accuracy with Rapid miner and WEKA data mining tools respectively. We have achieved best classification accuracy with Random Forest classifier in case of WEKA data mining tool for classification of banknote authentication. Table I shows that accuracy of classifiers with Rapid miner and WEKA data mining tools. Figure 2 shows that accuracy of classifiers with Rapid Miner and WEKA data mining tools.

Table I. Comparison of classifiers using Rapid Miner and WEKA data mining tools

Model	Rapid Miner	WEKA
Random Tree	55.60%	98.5%
Random Forest	57.54%	99.3%
Naive Bayes	51.68%	84.1%

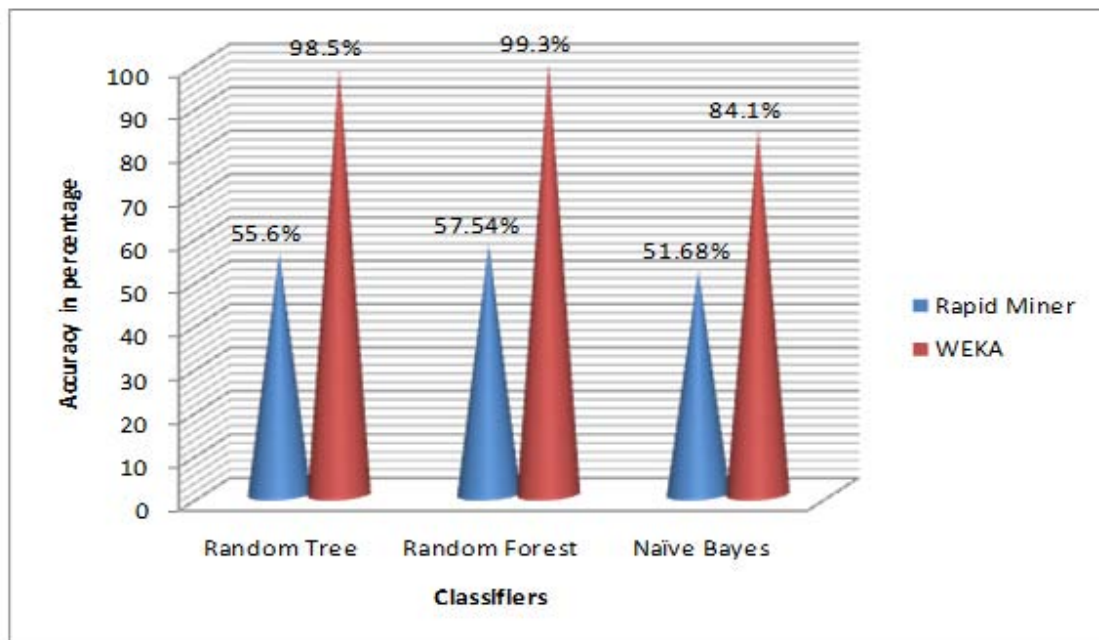


Figure 2. Accuracy of classifiers with Rapid Miner and WEKA data mining tools

In second section, compared the accuracy of classifiers using Rapid Miner and Tanagra data mining tools for classifying authentic and non-authentic notes. We have compared the classification accuracy of classifiers with Rapid Miner and Tanagra data mining tools. ID3 classifier gives 57.58% of accuracy with Rapid miner and K-NN

classifiers and gives 99.56% of accuracy with Tanagra data mining tool. The K-NN classifier gives better classifier with Tanagra data mining tool. Table II shows that accuracy of classifiers with Rapid miner and Tanagra data mining tools. Figure 3 shows that accuracy of classifiers with Rapid Miner and Tanagra data mining tools.

Table II . Comparison of classifiers using Rapid Miner and Tanagra data mining tools

Model	Rapid Miner	Tanagra
Random Tree	55.60%	98.32%
Naive Bayes	51.68%	85.25%
ID3	57.58%	90.29%
K-NN	54.88%	99.56%

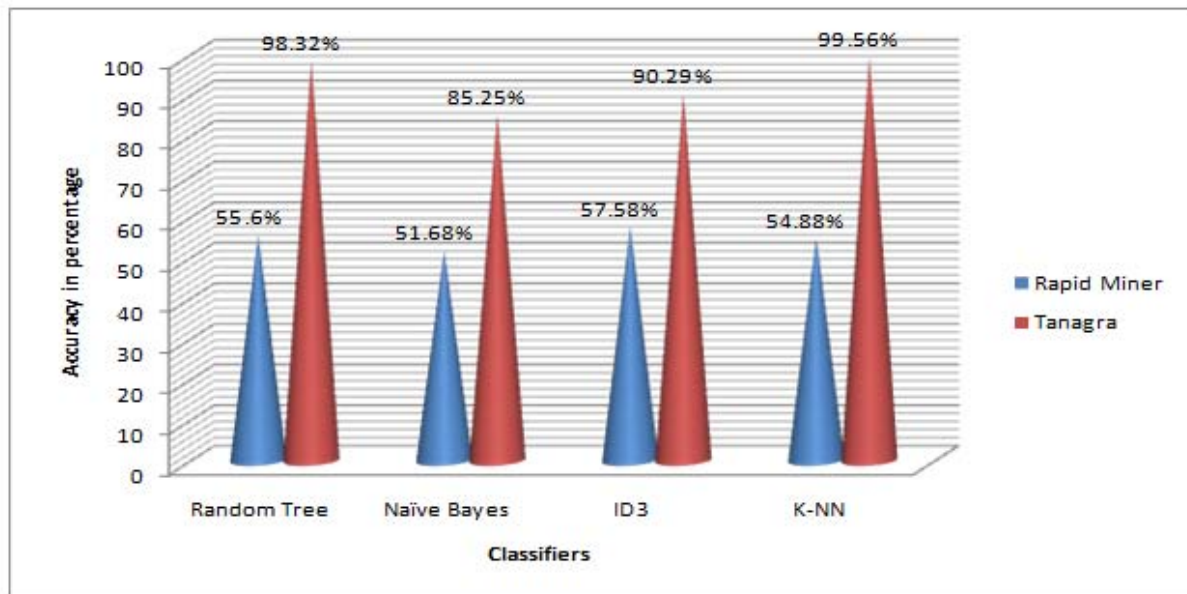


Figure 3. Accuracy of classifiers with Rapid Miner and Tanagra data mining tools

Finally, we have compared the performance of classifiers with WEKA and Tanagra data mining tools. MLP gives 98.83% of accuracy with WEKA and 98.61% of accuracy with Tanagra data mining tool. MLP achieved better classification accuracy with WEKA data mining tool for

classification of banknote authentication. Table III shows that accuracy of classifiers with WEKA and Tanagra data mining tools. Figure 4 shows that accuracy of classifiers with WEKA and Tanagra data mining tools.

Table III. Comparison of classifiers using WEKA and Tanagra

Model	WEKA	Tanagra
Random Tree	98.5%	98.32%
Naive Bayes	84.1%	85.25%
SVM	98.0%	98.10%
Multilayer Perceptron (MLP)	98.83%	98.61%

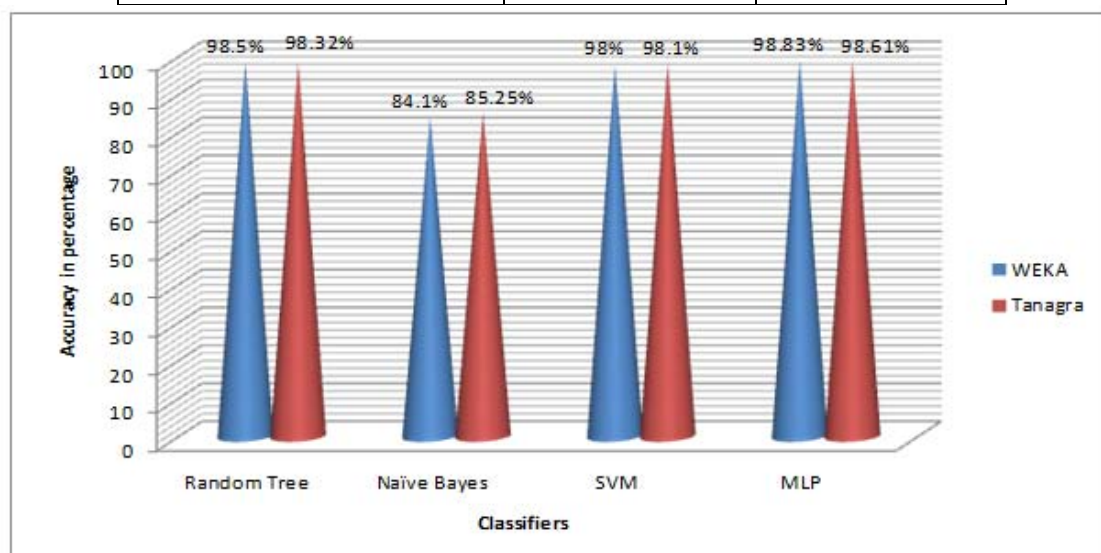


Figure 4. Accuracy of classifiers with WEKA and Tanagra data mining tools

We have analyzed the performance of common classifier available in WEKA, Tanagra and rapid Miner data mining tools. The Random tree gives better classification accuracy with WEKA data mining tool, similarly other common classifier is Naive Bays gives better classification accuracy with Tanagra data mining tool. Finally we have compared the overall comparison of classifier in which K-NN achieved 99.56% of accuracy with Tanagra data mining tool for banknote authentication.

IV. CONCLUSION

Classification is very important application of data mining technique to identify and classify the pattern. There are various researchers have worked in the field of classification of banknote authentication as authentic or not. This research work used various classification techniques with three different data mining tools like WEKA, Tanagra and Rapid Miner. First we have also compared the classification accuracy of classifiers with Rapid miner and WEKA data mining tools. Next we have compared the classification accuracy of classifiers with Rapid miner and Tanagra and lastly compared the classification accuracy of classifiers with WEKA and Tanagra data mining tools. Finally, concluded that K-NN is a robust classifier and achieved 99.56% of accuracy with Tanagra data mining tool for banknote authentication.

V. REFERENCE

- [1]. N. I. Nife , "Data mining techniques on Bank Note Authentication", International Journal of Engineering Science Invention, Vol. 5(2) , pp. 62-71, 2016.
- [2]. S. Jadhav and R. Khanai , "Authentication And Counterfeit Detection Of Currency Using Image Processing", International Conference on Recent Innovation in Engineering and Management, at Dahanjay Mahadik Group of Institutions Kolhapur, Maharashtra, held on 23 rd March 2016.
- [3]. C..R. Jyothi, Y.K. Sundara Krishna and V. S. Rao, "Bank Notes Authentication System based on Wavelet Features and Artificial Neural Network", International Journal of Computer Applications , Vol. 142(2), pp. 24-28, 2016.
- [4]. B. S. Prasanthi and R. Setty, "Indian Paper Currency Authentication System using Image Processing", International Journal of Scientific Research Engineering and Technology, Vol. 4(9), pp. 973-981, 2015.
- [5]. J. Han, M. Kamber and J. Pei , "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2016.
- [6]. S. Sen Sharma , "Bank Note Authentication: A Genetic Algorithm Supported Neural based Approach", International Journal of Advanced Research in Computer Science, Vol. 7(7), pp. 97-102, 2016.
- [7]. R. Ghosh and R. Khare , "A study on Diverse Recognition Technique for India Currency Note", International Journal of Engineering sciences and Research Technology, Vol. 2(6), pp. 1443-1447, 2013.
- [8]. S.E. Ali, M Gogoi and S. Mukherjee, "Challenges In Indian Denomination Recognition and Authentication", International Journal of Research in Engineering and Technology, Vol. 3, pp. 477-483, 2014.
- [9]. E. Kaya, A. Yasar and I. Saritas, "Banknote Classification Using Artificial Neural Network Approach", International Journal of Intelligent Systems and Applications in Engineering, Vol. 4(1), pp. 16-19, 2013.
- [10]. R. Mirza and V. Nanda, "Design and Implementation of Indian Paper Currency Authentication System Based on Feature Extraction by Edge Based Segmentation Using Sobel Operator", International Journal of Engineering Research and Development, Vol. 3(2), pp. 41-46, 2012.
- [11]. A. K. Pujari, "Data Mining Techniques", Universities Press (India) Private Limited, 4th ed., 2001.
- [12]. I. H. Witten. and , F. Eibe, "Data Mining Practical Machine Learning Tools and Techniques", Morgan Kaufmann, San, 2nd ed., 2005.
- [13]. Z., Tang, and J. MacLennan , "Data Mining with SQL Server 2005", Willey Publishing, Inc, USA, 2005.
- [14]. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml/datasets.html]. (Browsing date: 22-02-2017).
- [15]. WEKA Data Mining Tools: http://www.cs.waikato.ac.nz/~ml/weka/ (Browsing date: Feb 2017).
- [16]. Tanagra Data Mining Tool: http://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html. (Browsing date: Jan 2017).
- [17]. Rapid Miner Data Mining Tool: http://rapidminer.software.informer.com/download/#download_content (Browsing date: Feb 2017).