

Classifymeister End Evaluation

Team Members : Nitin Babu , Shikhar Singh , Mohit Dhakad , Chetan , Harsukh Sagri

Objective : Applications of Machine Learning in Detecting Afghan Fake Banknotes using Random Forest and Naïve Bayes algorithms.

Introduction :

- This research paper introduces a method that uses image processing to identify fake Afghan bank notes by analyzing the specific security features.
- The researchers extracted statistical features from images of the bank notes and use the WEKA machine learning tool to create models and classify the banknotes. They employed algorithms like **Random forest** and **Naïve Bayes** for classification.

Methodolgy :

- **Image acquisition :** High resolution images of bank notes were obtained using a scanner capturing both genuine and counterfeit banknotes.
- **Crop image from background :** Irrelevant background information in the acquired images was removed by isolating the largest blob using thresholding and masking techniques.
- **Pre – processing :** The banknote images underwent preprocessing steps such as grayscale conversion image adjustment noise removal and resizing to enhance feature extraction efficiency.
- **Crop Features :** Key features of the banknote images , specifically the Holographic strip and bottom design, were cropped as regions of interest(ROI).
- **Feature Extaraction :** Texture statistical-based feature extraction methods were utilized, focusing on first-order histogram-based features and second-order co-occurrence matrix features. These features capture texture and pattern information for distinguishing between genuine and counterfeit banknotes.
- **Dataset :** A dataset comprising both real and fake bankrupts was created based on the extracted texture features from the holographic strip and bottom design. The dataset comprises 17 attributes, where the first 16 attributes define the banknote characteristics and 17th one represent label “yes” or “no”.
- **Data processing :** Normalization was applied to scale all features and data resampling using synthetic minority oversampling technique(SMOTE) was performed to address class imbalance.
- **Classification :** Two machine learning algorithms(**Random forest** and **Naïve bayes**) over employed for binary classification of real and fake Afghan banknotes. The models were evaluated using 10- fold cross validation.

Algorithms :

Random Forest :

- Random forest is a supervised learning algorithm that is used for classification and regression tasks.
- It is an ensemble learning method that constructs multiple decision trees during training and makes predictions based on the mode of the classes or the mean prediction of the individual trees.
- The decision tree is a tree-like structure where non-leaf nodes represent feature tests, branches represent attribute values, and leaf nodes store categories.
- The decision process starts at the root node, tests feature attributes, selects branches based on attribute values, and continues until a leaf node is reached, which determines the final decision or prediction.

Naïve Bayes :

- Naive Bayes is a probabilistic machine learning algorithm based on Bayes' theorem with the assumption of independence between features.
- It is commonly used for classification tasks and is particularly suitable for text classification and spam filtering.
- The algorithm calculates the posterior probability of each class given the input features and predicts the class with the highest probability.
- Naive Bayes classifiers are relatively simple and fast to train and make predictions. They can handle large datasets with high dimensionality efficiently.

Evaluation :

- In this study, two classifiers (RF and NB) were evaluated on an imbalanced dataset of 1000 AFN banknotes. The application of SMOTE resampling technique significantly improved the performance of all classifiers, with RF achieving the highest accuracy of 99%, followed by NB with 97% .
- The RF classifier consistently outperformed the other classifier across various evaluation metrics such as **TP Rate, precision, recall, and F-measure**. SMOTE at different percentages (100%, 200%, and 300%) demonstrated its effectiveness in enhancing the performance of the classifiers, particularly RF, indicating the potential of this technique in improving detection of counterfeit banknotes.

Conclusion :

- The RF classifier exhibited the highest accuracy among the models tested. However, the study is limited to detecting counterfeit new 1000 AFN banknotes and may not be generalized to other denominations. Despite these limitations, their work represents an important step towards developing a reliable and cost-effective solution for detecting counterfeit Afghan banknotes.