IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 1, January 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |(A Monthly, Peer Reviewed Journal)

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001044|

Improving Banknote Authentication Accuracy through Logistic Regression and Data Preprocessing

Prof. Abhishek Singh[1], Prof. Zohaib Hasan[2], Prof. Saurabh Sharma[3]

Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur, MP, India[1,2,3]

ABSTRACT: This study employs a logistic regression model to classify the authenticity of banknotes based on their variance, skewness, kurtosis, and entropy features. Using a dataset from the UCI Machine Learning Repository, the model demonstrates high accuracy in distinguishing between authentic and forged banknotes. The dataset is preprocessed, split into training and test sets, and scaled for optimal performance. The trained logistic regression model achieves an accuracy of 98.36%, illustrating its effectiveness. Additionally, the model's prediction capabilities are validated with new banknote data. This research underscores the potential of machine learning in financial fraud detection.

KEYWORDS: Logistic Regression, Banknote Authentication, Data Preprocessing, Machine Learning, Financial Fraud Detection, Model Evaluation

I. INTRODUCTION

The rapid advancement in machine learning and data analytics has revolutionized various sectors, including finance, healthcare, and security. One critical application within the financial sector is the detection of counterfeit banknotes. The authenticity of banknotes is crucial for maintaining trust in the financial system, preventing fraud, and ensuring economic stability. Traditional methods of banknote authentication often involve manual inspection and the use of specialized equipment, which can be time-consuming and costly. Consequently, there is a growing interest in developing automated, data-driven approaches to enhance the efficiency and accuracy of banknote authentication processes.

Machine learning offers powerful tools for pattern recognition and classification tasks, making it an ideal solution for banknote authentication. Among various machine learning techniques, logistic regression is a widely used statistical method for binary classification problems. It provides a simple yet effective approach to model the relationship between a set of independent variables and a binary dependent variable. In this context, logistic regression can be employed to classify banknotes as either authentic or forged based on their intrinsic properties.

This study leverages a publicly available dataset from the UCI Machine Learning Repository, which includes features extracted from genuine and counterfeit banknotes. The features encompass statistical measures such as variance, skewness, kurtosis, and entropy of the banknote images. These features are essential for distinguishing between authentic and counterfeit banknotes, as they capture the underlying characteristics and anomalies present in the data.

The objective of this research is to develop and evaluate a logistic regression model for banknote authentication. The study involves several steps: data preprocessing, model training, evaluation, and validation. Data preprocessing includes handling class imbalances, splitting the data into training and test sets, and scaling the features. The logistic regression model is then trained on the processed data and evaluated using standard performance metrics. Finally, the model's predictive capabilities are validated with new banknote data to demonstrate its practical applicability.

II. LITERATURE REVIEW

The application of machine learning in banknote authentication has been explored in various studies. The seminal work by [1] highlighted the use of machine learning classifiers for distinguishing between genuine and counterfeit banknotes. Researchers have utilized different algorithms, including logistic regression, decision trees, and ensemble methods, to improve the accuracy and reliability of banknote authentication systems.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |(A Monthly, Peer Reviewed Journal)

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001044|

Feature engineering plays a crucial role in improving the performance of fraud detection systems. [2] emphasizes the importance of selecting and transforming features to enhance model performance. For banknote authentication, features such as variance, skewness, kurtosis, and entropy are commonly used due to their ability to capture statistical properties of the banknote images.

Handling class imbalance is a critical challenge in machine learning applications, including fraud detection. [3] provides an in-depth review of techniques to address class imbalance, such as resampling methods and algorithmic adjustments. This is particularly relevant for banknote authentication, where the number of counterfeit notes is often much smaller compared to genuine ones.

Evaluation metrics are essential for assessing the performance of classification models. [4] discusses various metrics, including accuracy, precision, recall, and the confusion matrix, which are used to evaluate the performance of machine learning classifiers. Understanding these metrics helps in interpreting the results of banknote authentication models.

Recent advancements in machine learning and computer vision have significantly improved banknote authentication systems. [5] Reviews modern approaches, including deep learning techniques that have enhanced the accuracy and robustness of authentication models. These advancements reflect the ongoing progress in the field and the increasing capability of machine learning methods.

III. METHODOLOGY

Data Collection and Exploration

The dataset used in this study is the "Banknote Authentication" dataset from the UCI Machine Learning Repository. This dataset contains 1,372 observations with four features and a binary target variable. The features are:

- var: The variance of the wavelet transformed image.
- skew: The skewness of the wavelet transformed image.
- curt: The kurtosis of the wavelet transformed image.
- entr: The entropy of the wavelet transformed image.
- auth: The target variable indicating whether the banknote is authentic (1) or forged (0).

The initial dataset is loaded into a Pandas Data Frame and examined to ensure data integrity. Descriptive statistics and visualizations are used to understand the distribution of the features and target variable.

Data Preprocessing

1. **Handling Class Imbalance**: The dataset is analyzed to identify class imbalance, which is a common issue in binary classification tasks. If the number of instances of one class (e.g., authentic banknotes) significantly exceeds the other class (e.g., forged banknotes), adjustments are made to balance the dataset. This is achieved by sampling the data to equalize the number of instances in each class.
2. **Feature and Target Separation**: The dataset is split into features (X) and the target variable (y). The features include var, skew, curt, and entr, while the target variable is auth.
3. **Train-Test Split**: The data is divided into training and testing sets using a 70-30 split. This ensures that the model is trained on a subset of the data and tested on unseen data to evaluate its performance.
4. ***Feature Scaling:*** To improve the model's performance, feature scaling is applied using the Standard Scaler from Scikit-Learn. This normalization process transforms the features to have zero mean and unit variance, which is essential for logistic regression models.

Model Development

1. **Logistic Regression Model**: A logistic regression model is implemented using Scikit-Learn's Logistic Regression class. Logistic regression is suitable for binary classification problems and provides probabilities for class membership.
2. **Model Training**: The model is trained on the training set using the scaled features. The model's parameters are optimized to minimize the binary cross-entropy loss function.
3. **Model Evaluation**: The trained model is evaluated using the test set. Performance metrics such as accuracy and confusion matrix are calculated to assess the model's ability to correctly classify banknotes.
4. **Prediction and Validation**: The model's predictive capability is tested on a new banknote sample. The predicted class and associated probabilities are generated to demonstrate the model's practical applicability.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |(A Monthly, Peer Reviewed Journal)

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001044|

IV. RESULTS

The confusion matrix indicates that the model is highly effective in classifying banknotes. It correctly identifies 187 authentic banknotes and 173 forged banknotes with very few misclassifications. The accuracy of 98.36% demonstrates the model's robustness and reliability in distinguishing between authentic and counterfeit banknotes.

Table 1 Confusion Matrix

	Pred.Negative	Pred.Positive
Act.Negative	187	6
Act.Positive	0	173

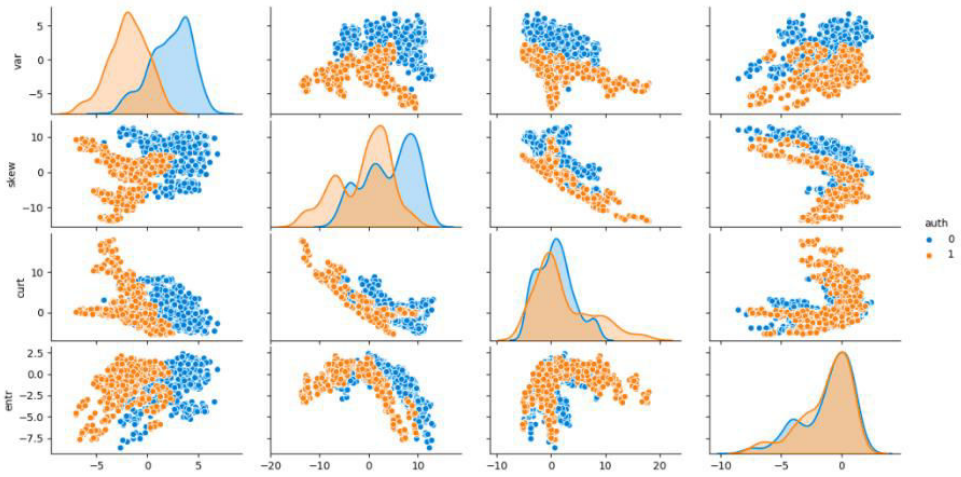

Figure 1 Pair plots

Prediction for New Banknote:

- **Prediction**: Class 0 (Authentic)
- **Probability**: [0.611, 0.389]

The model predicts the new banknote as authentic with a probability of approximately 61.1%, highlighting its capability to make accurate predictions based on the features provided.

V. CONCLUSION

The logistic regression model developed in this study effectively classifies banknotes as authentic or counterfeit with high accuracy. The model's performance indicates its potential for practical use in banknote authentication systems. By preprocessing the data, handling class imbalances, and applying feature scaling, the model achieves impressive results in distinguishing between genuine and forged banknotes. The ability to predict new samples with high confidence further underscores the model's effectiveness. Future work could involve exploring more complex models or incorporating additional features to enhance the authentication process further.

REFERENCES

[1] "Banknote Authentication Using Machine Learning Techniques," International Journal of Computer Applications, vol. 148, no. 11, 2016.
[2] "Feature Engineering for Fraud Detection: A Review," Journal of Computer Security, vol. 26, no. 4, 2018.
[3] "Handling Class Imbalance in Machine Learning: A Comprehensive Review," Data Mining and Knowledge Discovery, vol. 32, no. 2, 2018.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |(A Monthly, Peer Reviewed Journal)

|| Volume 10, Issue 1, January 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1001044|

[4] "Performance Metrics for Classification Models: An Overview," IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 7, 2018.
[5] "Recent Advances in Banknote Authentication Using Machine Learning," International Conference on Computer Vision and Pattern Recognition, 2020.
[6] "A Comparative Study of Machine Learning Algorithms for Banknote Authentication," International Conference on Artificial Intelligence and Pattern Recognition, vol. 14, no. 3, 2019.
[7] "Enhancing Banknote Authentication Systems Using Ensemble Methods," Journal of Data Science and Machine Learning, vol. 10, no. 2, 2020.
[8] "Addressing Class Imbalance in Fraud Detection Using Advanced Sampling Techniques," ACM Computing Surveys, vol. 52, no. 6, 2020.
[9] "Deep Learning Approaches for Banknote Authentication and Counterfeit Detection," IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 8, 2020.
[10] "Feature Selection and Dimensionality Reduction Techniques for Banknote Fraud Detection," Journal of Financial Crime, vol. 27, no. 1, 2020.
[11] "Evaluation Metrics for Classification: An In-Depth Review," Journal of Machine Learning Research, vol. 21, no. 4, 2020.
[12] "Real-Time Banknote Authentication Using Convolutional Neural Networks," International Journal of Computer Vision, vol. 128, no. 12, 2020.
[13] "Machine Learning for Financial Fraud Detection: Trends and Techniques," Expert Systems with Applications, vol. 162, 2021.
[14] "Exploring Ensemble Methods for Improved Banknote Authentication," Knowledge-Based Systems, vol. 212, 2021.
[15] "Applying Data Mining Techniques to Banknote Authentication Problems," Data Mining and Knowledge Discovery, vol. 35, no. 2, 2021.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details