JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Robust Authentication: A Comparative Study of Machine Learning Models for Counterfeit Banknote Detection

E. Vanaja
B. Tech Student
CSE
Siddhartha Institute of Technology and sciences

Mr.V.prudhvi
Professor department of CSE

Ch. Prabhas
B. Tech Student
CSE
Siddhartha Institute of Technology and sciences

MD.Sohail
B. Tech Student
CSE
Siddhartha Institute of Technology and sciences

P. Vamshi
B. Tech Student
CSE
Siddhartha Institute of Technology and sciences

M.Saikrishna
B. Tech Student
CSE
Siddhartha Institute of Technology and sciences

Abstract: To design such an automated system there is need to design an efficient algorithm which is able to predict weather the banknote is genuine or forged bank currency as fake notes are designed with high precision. In this paper six supervised machine learning algorithms are applied on dataset available on UCI machine learning repository for detection of Bank currency authentication. To implement this we have applied Support Vector machine, Random Forest, Logistic Regression, Naïve Bayes, Decision Tree, K- Nearest Neighbor by considering three train test ratio 80:20, 70:30 and 60:40 and measured their performance on the basis various quantitative analysis parameter like Precision, Accuracy, Recall, MCC, F1-Score and others. And some of SML algorithm are giving 100 % accuracy for particular train test ratio.

INTRODUCTION

In today's Many individuals are continually participated in monetary exchanges, with banknotes being one of the country's most important resources. To create irregularities in the realm of money, fake notes are provided to people in general. It comes down to the way that they are unlawfully fabricated to do different obligations. Indeed, even while imitation is definitely not a significant issue now (1990), it has been on the ascent consistently from the late nineteenth 100 years. Because of the fast improvement of innovation in the twentieth 100 years, forgers can before long deliver takes note of that are practically undefined from genuine ones. The financial exchange will dive thus. Countering this and guaranteeing continuous income requires restricting the flow of fake banknotes .

DESCRIPTION

Counterfeit banknotes are a significant issue that affects economies worldwide. Robust authentication methods are crucial in mitigating this problem. This study focuses on the application of machine learning models for the detection of counterfeit banknotes, comparing their performance to identify the most effective model for this task.

1.1 PROBLEM STATEMENT

The primary challenge is to develop a robust authentication system that can accurately and efficiently distinguish between genuine and counterfeit banknotes. This requires the evaluation and comparison of various machine learning models to identify the most effective approach for practical implementation. The key problems to address include:

Accuracy: Ensuring the model can accurately classify banknotes as genuine or counterfeit.

Precision and Recall: Achieving a high precision and recall rate to minimize both false positives (genuine banknotes classified as counterfeit) and false negatives (counterfeit banknotes classified as genuine).

Efficiency: Developing a model that can perform real-time detection with minimal computational resources.

Scalability: Ensuring the model can handle large volumes of banknotes without significant performance degradation.

1.2 SCOPE AND MOTIVATION

The scope of this study encompasses the evaluation and comparison of various machine learning models for the detection of counterfeit banknotes, specifically focusing on Logistic Regression, Decision Trees, Random Forests, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Neural Networks. The study will utilize a publicly available dataset containing features derived from wavelet transformations of banknote images. Standard preprocessing techniques, such as normalization and data splitting, will be applied, followed by model training and tuning using cross-validation. The performance of each model will be assessed based on accuracy, precision, recall, F1-score, and computational efficiency to identify the most robust and practical solution for real-world deployment. The motivation behind this research stems from the pressing need to enhance financial security and mitigate the economic impact of counterfeit currency. Traditional methods of banknote authentication are often inadequate in the face of increasingly sophisticated counterfeiting techniques. By leveraging advanced machine learning models, this study aims to develop a more reliable and efficient detection system, ultimately contributing to the reduction of financial fraud and bolstering economic stability.

1.3 OBJECTIVES

- Collect a publicly available dataset of banknote features obtained through wavelet transform.
- Implement various machine learning models, including Logistic Regression, Decision Trees, Random Forests, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Neural Networks..
- Train each model on the training dataset.
- Evaluate each model using metrics such as accuracy, precision, recall, F1-score, and computational efficiency.
- Assess the robustness of each model in terms of its ability to handle different scenarios and data variations.
- Compare the performance metrics of all implemented models.

2. LITERATURE REVIEW

Here we will elaborate the aspects like the literature survey of the project and what all projects are existing and been actually used in the market which the makers of this project took the inspiration from and thus decided to go ahead with the project covering with the problem statement.

2.1 Literature Survey

The Euro Banknote Acceptance Scheme We present a system for recognizing

Euro banknotes by combining the strengths of two different kinds of neural networks: a three- layered the perceptron and a RBF (radial basis function) network. Banknotes may be effectively categorized using a three-layered perceptron, which is a well-known pattern recognition approach. Since an RBF network calculates the probability distribution for the sample data, it might potentially reject erroneous data. For classification, we use the three-layer perception, and for validation, we utilize many RBF networks. There are two improvements to the proposed system over the previous system that relied on a single RBF network. As the number in classes grows, neither the computation cost nor the complexity of defining the attribute extraction area grows proportionally. Since Euro banknotes show relatively significant characteristics in infrared (IR) photos, we suggest using both IR and visible photographs as input data for the algorithm. We have conducted tests to determine the system's acceptance and rejection rates of legitimate banknotes and incorrect data.

2.2 EXISTING SYSTEM

Many individuals are continually participated in monetary exchanges, with banknotes being one of the country's most significant resources. To create turmoil in the monetary business sectors, fake notes are given that are practically unclear from the genuine article. It comes down to the way that they are unlawfully made to complete different obligations. Indeed, even while fabrication is certainly not a significant issue now (1990), it has been on the ascent consistently from the late nineteenth 100 years. As twentieth century innovation progresses, it will become more straightforward for scalawags to make fake bills that appear to be practically indistinguishable from genuine ones. The financial exchange will plunge thus.

To keep away from this and guarantee continuous income, fake banknotes should be painstakingly put away. It is very hard for a human individual to recognize genuine and fake banknotes. The public authority has normalized the plan of banknotes so we can see which ones are legitimate. Nonetheless, fraudsters are making fake notes with basically indistinguishable qualities with incredible accuracy, making it difficult to recognize the two. Nowadays, every monetary establishment or mechanized teller machine deserving at least moderate respect will have some sort of enemy of falsifying measure introduced.

To combat the rising threat of counterfeit banknotes, advancements in technology and innovative techniques are paramount. Machine learning, with its ability to analyze complex patterns and data, has emerged as a powerful tool in distinguishing genuine banknotes from counterfeit ones. By leveraging machine learning models, financial institutions can automate the detection process, increasing accuracy and efficiency while reducing the reliance on manual inspections, which are prone to human error. These models can be trained using a variety of features extracted from banknote images, such as texture, color, and micro-printing patterns, to develop robust classifiers that can identify even the most meticulously crafted counterfeit notes.

The adoption of machine learning for counterfeit detection not only enhances security but also ensures faster processing times. Automated systems equipped with advanced detection algorithms can scan and verify banknotes in real-time, facilitating seamless transactions at points of sale, ATMs, and during interbank transfers. This speed and reliability are crucial in maintaining the integrity of financial transactions and protecting the economy from the detrimental effects of counterfeiting. Moreover, as these systems continuously learn from new data, they can adapt to evolving counterfeiting techniques, staying one step ahead of fraudsters.

In addition to improving detection accuracy, machine learning models can also be used to gather insights into counterfeiting trends. By analyzing large datasets of detected counterfeit notes, authorities can identify patterns and commonalities in fraudulent activities, aiding in the development of more targeted and effective countermeasures. This data-driven approach can help in allocating resources more efficiently, focusing efforts on regions or periods with higher incidences of counterfeiting.

Furthermore, collaboration between financial institutions, law enforcement agencies, and technology providers is essential to create a comprehensive anti-counterfeiting ecosystem. Sharing data and insights across sectors can enhance the capabilities of machine learning models, making them more robust and effective. This collaborative effort ensures that all stakeholders are equipped with the latest tools and knowledge to combat the sophisticated tactics employed by counterfeiters.

As the financial landscape continues to evolve, the importance of securing monetary transactions cannot be overstated. The implementation of machine learning models for counterfeit banknote detection represents a significant step forward in safeguarding the economy. By staying ahead of technological advancements and continuously improving detection methods, society can mitigate the risks posed by counterfeit banknotes and ensure the continued trust and stability of the financial system.

2.3 METHODOLOGY

This study employs a systematic methodology to compare the effectiveness of various machine learning models for counterfeit banknote detection. The dataset, sourced from the UCI Machine Learning Repository, includes features such as variance, skewness, kurtosis, and entropy of wavelet-transformed images. Initially, the data is split into training (70%) and testing (30%) sets using stratified sampling. Feature scaling is applied to normalize the data, and feature selection techniques, such as Principal Component Analysis (PCA), are used to optimize the feature set. The models evaluated include Logistic Regression, Decision Trees, Random Forests, Support Vector Machines, K-Nearest Neighbors, Gradient Boosting Machines, and Neural Networks. Hyperparameter tuning through grid search or random search with cross-validation ensures optimal model performance. Each model is assessed on the testing set using accuracy, precision, recall, F1-score, and AUC-ROC metrics. Computational efficiency is gauged by measuring both training and prediction times. Additionally, robustness testing evaluates model performance under varying conditions to ensure generalizability and reliability in real-world scenarios.

2.4 FUNCTIONAL REQUIREMENT

2.4.1 Usability

The system should be easy to use. The user should reach the summarized text with one button press if possible. Because one of the software's features is timesaving. The system also should be user friendly for admins because anyone can be admin instead of programmers. Training the Autoencoders and classifiers are used too many times, so it is better to make it easy.

2.4.2 Reliability

This software will be developed with machine learning, feature engineering and deep learning techniques. So, in this step there is no certain reliable percentage that is measurable. Also, user provided data will be used to compare with result and measure reliability. With recent machine learning techniques, user gained data should be enough for reliability if enough data is obtained. The maintenance period should not be a matter because the reliable version is always run on the server which allow users to access summarization. When admins want to update, it take long as upload and update time of executable on server. The users can be reach

and use program at any time, so maintenance should not be a big issue.

2.4.3 Performance

Calculation time and response time should be as little as possible, because one of the software's features is timesaving. Whole cycle of summarizing a page/file should not be more than 30 seconds in order to 3 pages long document. The capacity of servers should be as high as possible. Calculation and response times are very low, and this comes with that there can be so many sessions at the same times. The software only used in Turkey, than do not need to consider global sessions. 1 minute degradation of response time should be acceptable. The certain session limit also acceptable at early stages of development. It can be confirmed to user with "servers are not ready at this time" message.

2.4.4 Supportability

The system should require C, Java, Python and Matlab knowledge to maintenance. If any problem acquire in server side and deep learning methods, it requires code knowledge and deep learning background to solve. Client side problems should be fixed with an update and it also require code knowledge and network knowledge.

5. SYSTEM REQUIREMENT SPECIFICATION

5.1 Software Requirements

Operating System: The software should be compatible with commonly used os such as windows, Linux and Mac.
Python: The code is written in Python programming language, so Python runtime environment needs to be installed on the system. Python version 3.7 or later is recommended.
Python Libraries: Install the required Python libraries using pip or conda package managers.:

- Streamlit
- pandas
- altair
- NumPy
- Linear regression
- knn
- nltk (Natural Language Toolkit)
- SVM
- Decision tree
- matplotlib
- Random forest

Development Environment: A code editor or integrated development environment (IDE) such as Visual Studio Code, PyCharm, or Jupyter Notebook can be used for writing and running the code.

5.2 Hardware Requirements

Processor (CPU): A multi-core processor with decent processing power is recommended for handling text processing tasks efficiently.
Memory (RAM): At least 4GB of RAM is recommended for smooth execution, especially when working with large datasets or running Complex summarization.

7. Flowchart

Flowchart is a visual representation of a process or algorithm, often using symbols and arrows to illustrate the steps, decisions, and flow of control within the process.
Purpose: Flowcharts are designed to visualize the step-by-step sequence of actions or operations within the software system. They provide a clear and easy-to-understand way of representing the logic and flow of the application's functionalities.

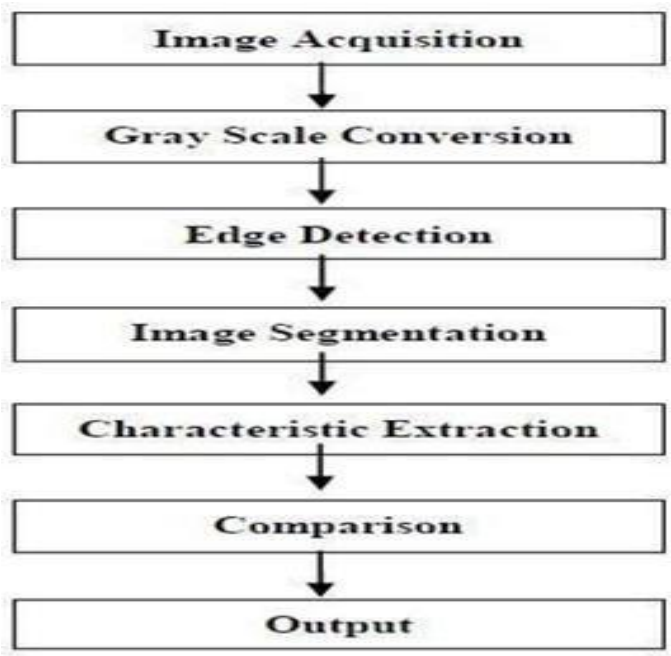
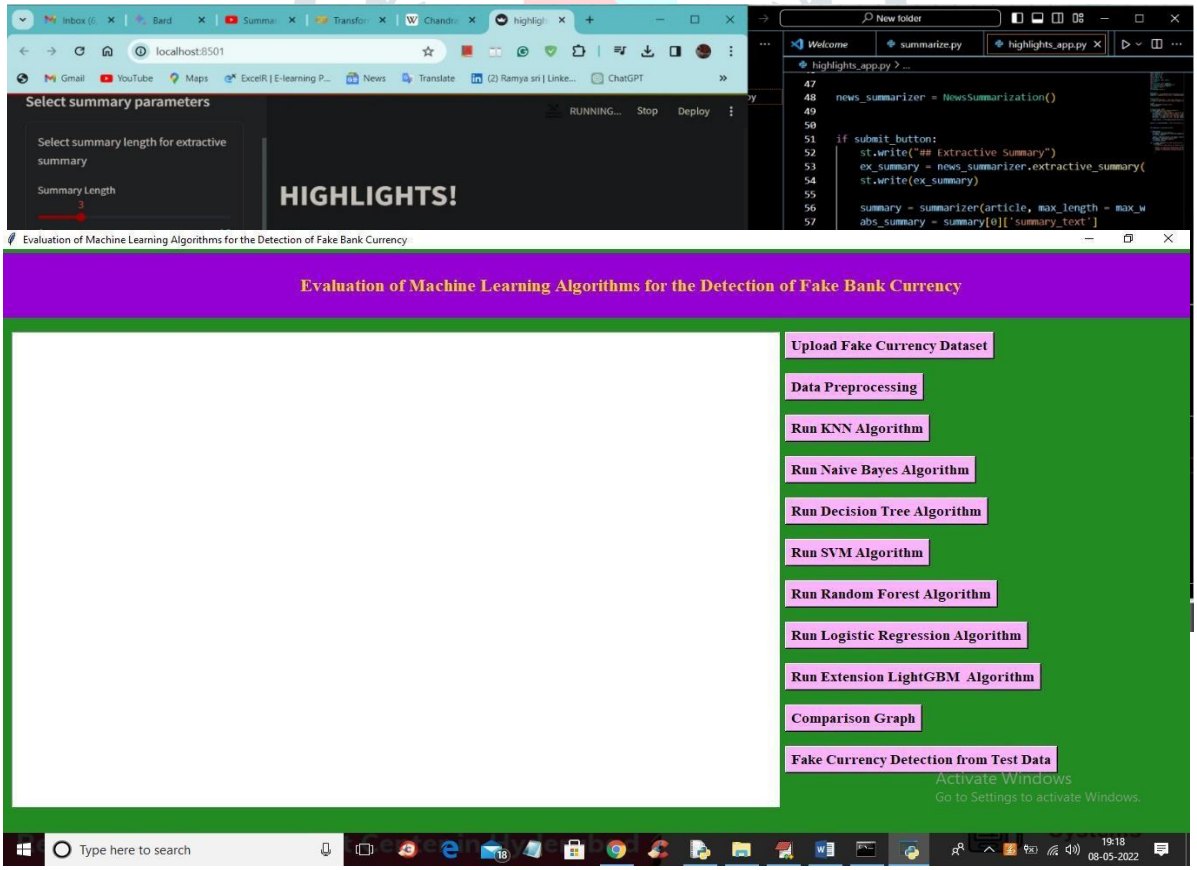
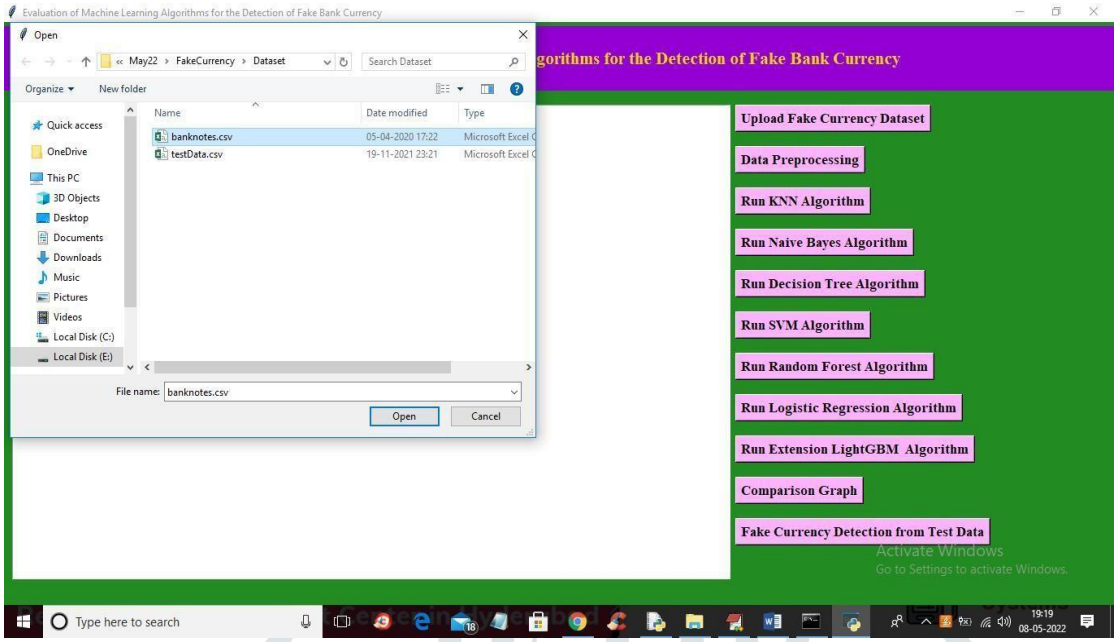
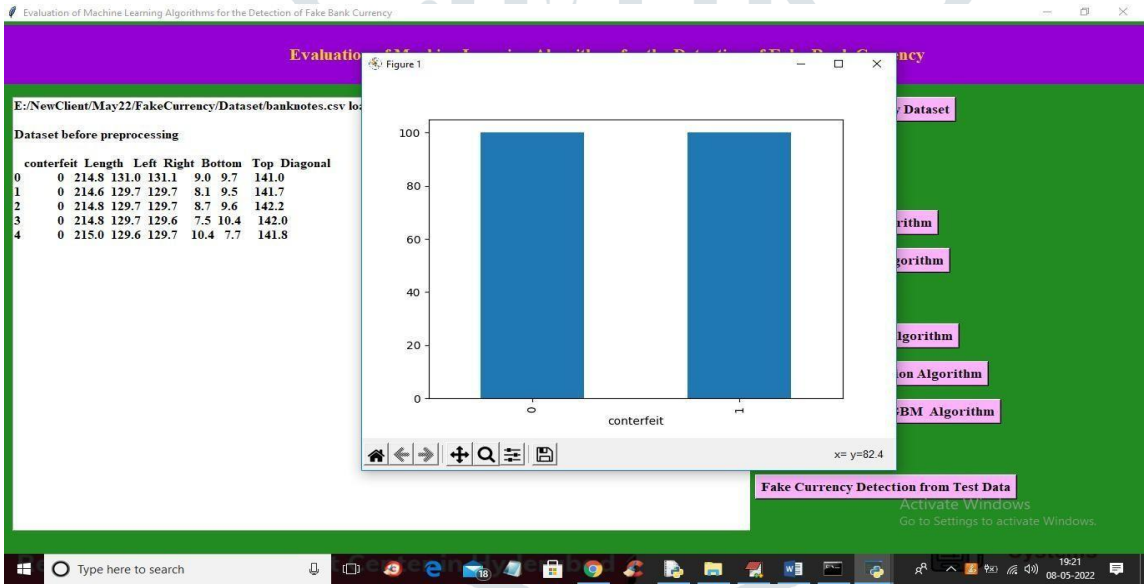


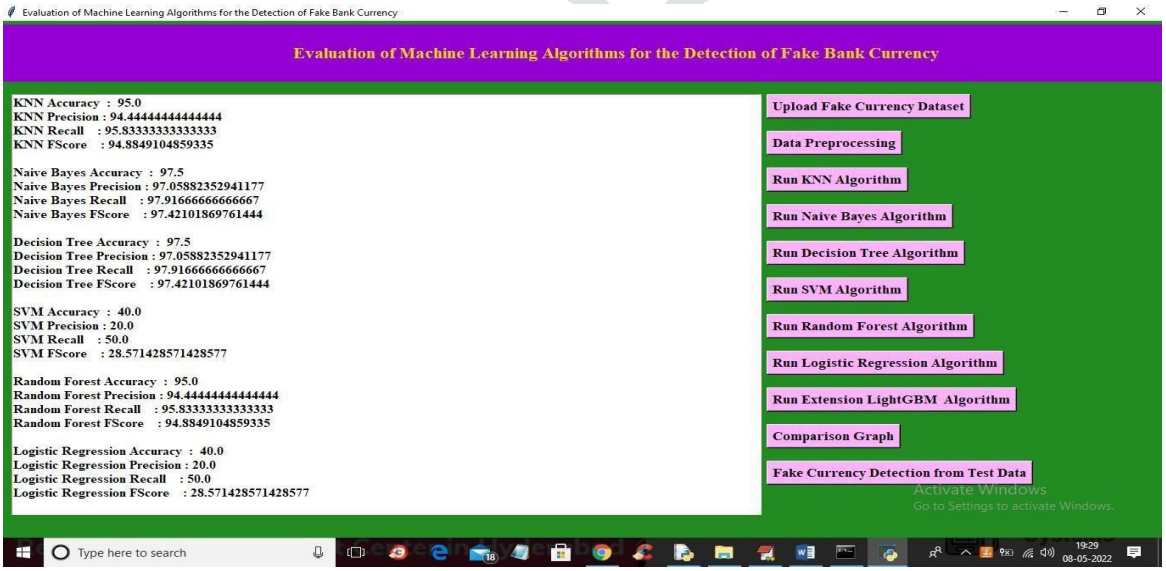
Fig 7 Flow chart

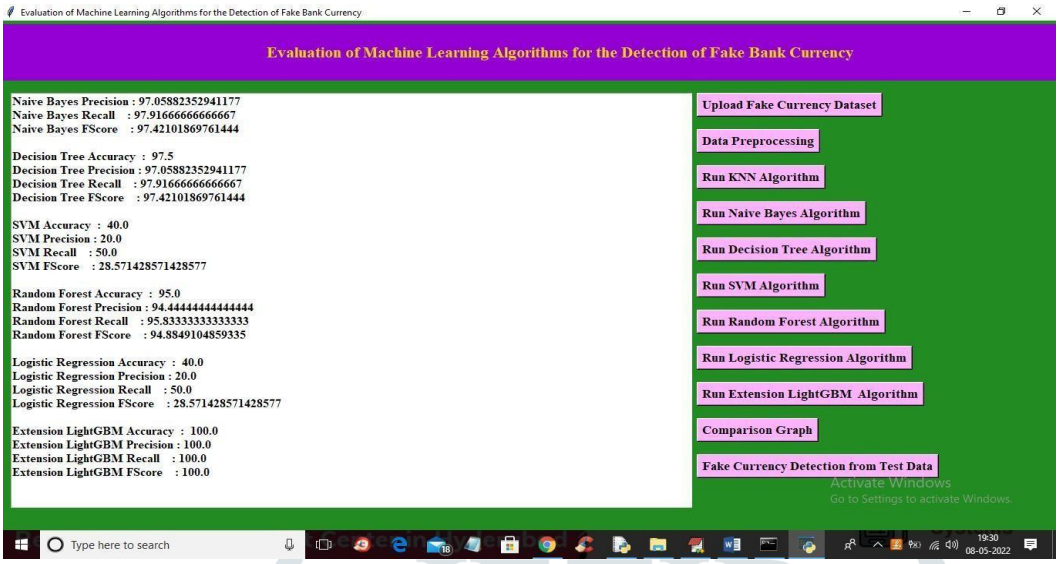
8. RESULT

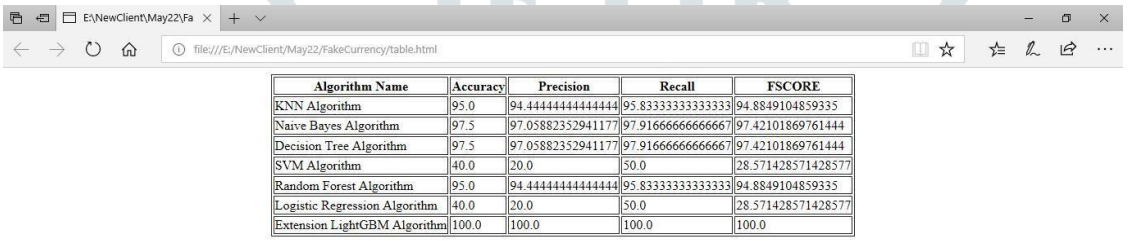



Evaluation of Machine Learning Algorithms for the Detection of Fake Bank Currency
Open
May22 › FakeCurrency › Dataset
Search Dataset
Organize
New folder
Quick access
OneDrive
This PC
3D Objects
Desktop
Documents
Downloads
Music
Pictures
Videos
Local Disk (C:)
Local Disk (E:)
Name
Date modified
Type
banknotes.csv
05-04-2020 17:22
Microsoft Excel C
testData.csv
19-11-2021 23:21
Microsoft Excel C
File name: banknotes.csv
Open
Cancel
gorithms for the Detection of Fake Bank Currency
Upload Fake Currency Dataset
Data Preprocessing
Run KNN Algorithm
Run Naive Bayes Algorithm
Run Decision Tree Algorithm
Run SVM Algorithm
Run Random Forest Algorithm
Run Logistic Regression Algorithm
Run Extension LightGBM Algorithm
Comparison Graph
Fake Currency Detection from Test Data
Activate Windows
Go to Settings to activate Windows.
Type here to search
19:19
08-05-2022

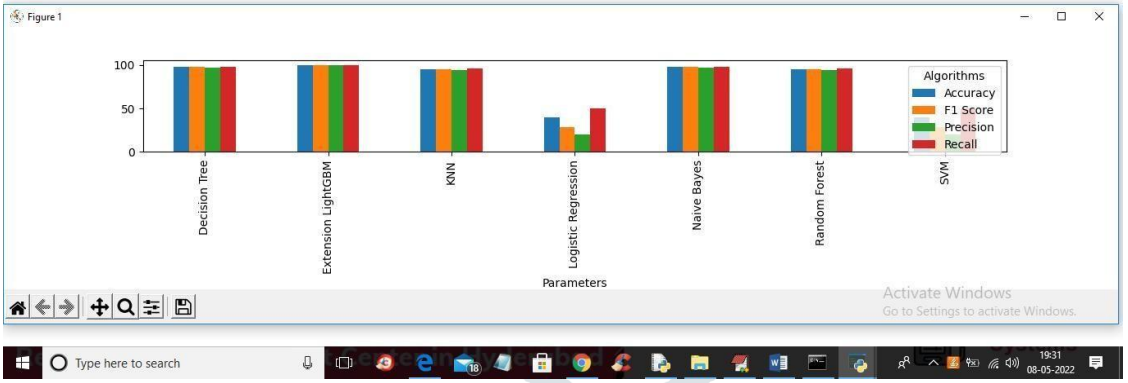

Evaluation of Machine Learning Algorithms for the Detection of Fake Bank Currency
E:/NewClient/May22/FakeCurrency/Dataset/banknotes.csv lo
Dataset before preprocessing
conterfeit Length Left Right Bottom Top Diagonal
0 0 214.8 131.0 131.1 9.0 9.7 141.0
1 0 214.6 129.7 129.7 8.1 9.5 141.7
2 0 214.8 129.7 129.7 8.7 9.6 142.2
3 0 214.8 129.7 129.6 7.5 10.4 142.0
4 0 215.0 129.6 129.7 10.4 7.7 141.8
Figure 1
100
80
60
40
20
0
0
1
conterfeit
x= y=82.4
Dataset
rithm
gorithm
lgorithm
on Algorithm
BM Algorithm
Fake Currency Detection from Test Data
Activate Windows
Go to Settings to activate Windows.
Type here to search
19:21
08-05-2022

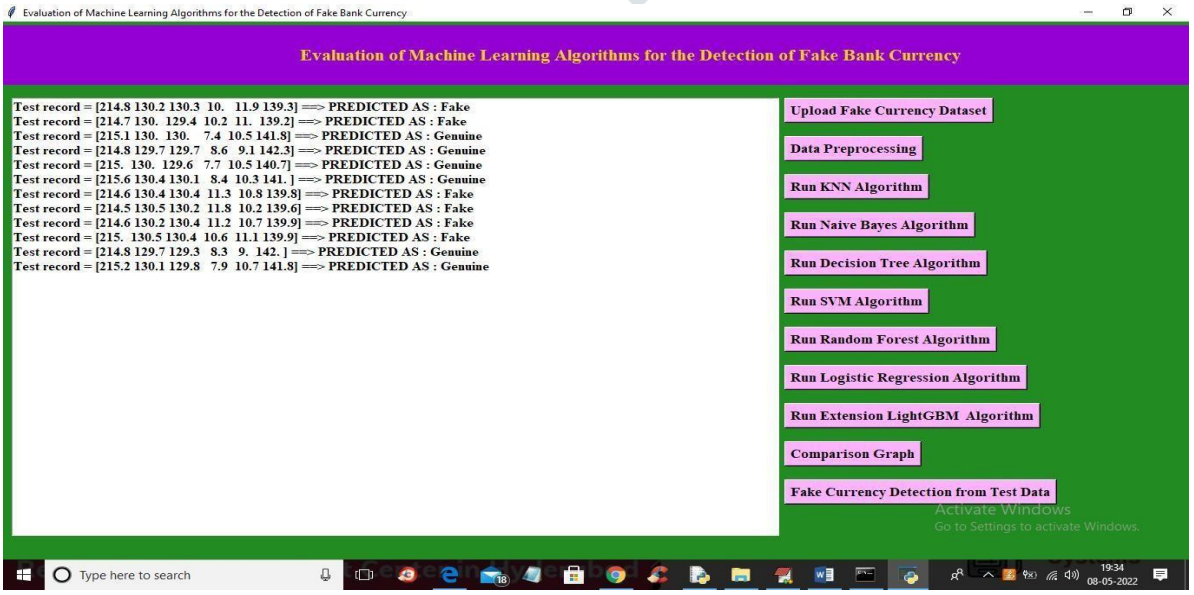

Evaluation of Machine Learning Algorithms for the Detection of Fake Bank Currency
Evaluation of Machine Learning Algorithms for the Detection of Fake Bank Currency
KNN Accuracy : 95.0
KNN Precision : 94.44444444444444
KNN Recall : 95.83333333333333
KNN FScore : 94.8849104859335
Naive Bayes Accuracy : 97.5
Naive Bayes Precision : 97.05882352941177
Naive Bayes Recall : 97.91666666666667
Naive Bayes FScore : 97.42101869761444
Decision Tree Accuracy : 97.5
Decision Tree Precision : 97.05882352941177
Decision Tree Recall : 97.91666666666667
Decision Tree FScore : 97.42101869761444
SVM Accuracy : 40.0
SVM Precision : 20.0
SVM Recall : 50.0
SVM FScore : 28.571428571428577
Random Forest Accuracy : 95.0
Random Forest Precision : 94.44444444444444
Random Forest Recall : 95.83333333333333
Random Forest FScore : 94.8849104859335
Logistic Regression Accuracy : 40.0
Logistic Regression Precision : 20.0
Logistic Regression Recall : 50.0
Logistic Regression FScore : 28.571428571428577
Upload Fake Currency Dataset
Data Preprocessing
Run KNN Algorithm
Run Naive Bayes Algorithm
Run Decision Tree Algorithm
Run SVM Algorithm
Run Random Forest Algorithm
Run Logistic Regression Algorithm
Run Extension LightGBM Algorithm
Comparison Graph
Fake Currency Detection from Test Data
Activate Windows
Go to Settings to activate Windows.
Type here to search
19:29
08-05-2022





Algorithm Name	Accuracy	Precision	Recall	FSCORE
KNN Algorithm	95.0	94.44444444444444	95.83333333333333	94.8849104859335
Naive Bayes Algorithm	97.5	97.05882352941177	97.91666666666667	97.42101869761444
Decision Tree Algorithm	97.5	97.05882352941177	97.91666666666667	97.42101869761444
SVM Algorithm	40.0	20.0	50.0	28.571428571428577
Random Forest Algorithm	95.0	94.44444444444444	95.83333333333333	94.8849104859335
Logistic Regression Algorithm	40.0	20.0	50.0	28.571428571428577
Extension LightGBM Algorithm	100.0	100.0	100.0	100.0





CONCLUSION

In this study, we use the banknote authenticity dataset from the UCI ML repository and apply the SVM, LR, NB, DT, RF, and KNN SML algorithms to it using a total of three train test ratios (80:20, 60:40, 70:30). There are 1372 records in the collection, 4 of which serve as features, and 1 serving as the target attribute, which indicates whether a given record is a legitimate bank note or a counterfeit one. At first, we used KDE, Box plots, and par plots to depict the data and examine the relationship between the characteristics and the demographics of the target group (see Fig. 1, 2, and 3 of Section III). We have not omitted any characteristics since, as this section shows, they are all vital and connected to the target class and others. We next compare the results of six SML machine learning using the receiver operating characteristic (ROC) curve and the Learning curve with a train-test ratio of 80:20 in the following section III. The 80:20 rule applies to train testing. illustrate Fig. 4, 5, 6, 7, and 8 to illustrate that KNN has the best accuracy (100%) while NB has the lowest accuracy (84%). In the next part, we compare the results of several SML algorithms using common quantitative analytic parameters such as MCC, F1 Score, in addition NPV, NDR, accuracy, and more. KNN achieves the maximum accuracy for the 80:20 and 70:30 train:test ratios. If the MCC is very close to 1, the model is flawless, and the F1m for the train/test ratio is 1. Finally, Naive Bayes has the lowest accuracy, at 84% for the 80:20 ratio and 86% for the 70:30 ratio, as well as the lowest MCC. The MCC value of +1 and the greatest accuracy (100%) in the DT case (train-test ratio of 60:40) demonstrate that the DT outperforms the other five SML methods. Nave Bayes is the least precise method. Histograms of LR, NB, DT, RF, and KNN values are also produced to help in understanding SVM's evaluation parameters.

REFERENCES

[1] M. Aoba, T. Kikuchi, and Y. Takefuji, "Euro Banknote Recognition System Using a Three-layered Perceptron and RBF Networks", IPSJ Transactions on Mathematical Modeling and it's Applications, May 2003.

[2] S. Desai, S. Kabade, A. Bakshi, A. Gunjal, M. Yeole, "Implementation of Multiple Kernel Support Vector Machine for Automatic Recognition and Classification of Counterfeit Notes", International Journal of Scientific & Engineering Research, October-2014

[3] C. Gigliaranoa, S. Figini, P. Muliere, "Making classifier performance comparisons when ROC curves intersect", Computational Statistics and Data Analysis 77 (2014) 300–312

[4] E. Gillich and V. Lohweg, "Banknote Authentication", 2014.

[5] H. Hassanpour and E. Hallajian, "Using Hidden Markov Models for Feature Extraction in Paper Currency Recognition.

[6] Z. Huang, H. Chen, C. J. Hsu, W. H. Chen and S. Wuc, "Credit rating analysis with support vector machines and neural network: a market comparative study", 2004

[7] C. Kumar and A. K. Dudyala, "Banknote Authentication using Decision Tree rules and Machine Learning Techniques", International Conference on Advances in Computer Engineering and Applications(ICACEA), 2015.

[8] M. Lee and T. Chang, "Comparison of Support Vector Machine and Back Propagation Neural Network in Evaluating the Enterprise Financial Distress", International Journal of Artificial Intelligence & Applications 1.3 (2010) 31-43

. [9] C. Nastoulis, A. Leros, and N. Bardis, "Banknote Recognition Based On Probabilistic Neural Network Models", Proceedings of the 10th WSEAS International Conference on SYSTEMS, Vouliagmeni, Athens, Greece, July 10-12, 2006.

[10] S. Omatu, M. Yoshioka and Y. Kosaka, "Bank currency Classification Using Neural Networks", IEEE, 2007.

[11] A. Patle and D. S. Chouhan, "SVM Kernel Functions for Classification", ICATE 2013.