

Interpretable AI-Enhanced Machine Learning for Intelligent Spambot and Fake Followers Identification

Mattada Venuka¹, Dr. A. Krishna Mohan²

¹PG Scholar, Department of Computer Science and Engineering, UCEK, JNTUK, Kakinada, Andhra Pradesh, India.

²Professor, Department of Computer Science and Engineering, UCEK, JNTUK, Kakinada, Andhra Pradesh, India.

How to cite this paper:

Mattada Venuka¹, Dr. A. Krishna Mohan²,
"Interpretable AI-Enhanced Machine
Learning for Intelligent Spambot and Fake
Followers Identification", IJIRE-V7I4-246-
257.



Copyright © 2026
by author(s) and
Fifth Dimension
Research

Publication. This work is licensed under the
Creative Commons Attribution International
License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>

Abstract: Social spambots are automated accounts that imitate normal users and can be used to generate misleading, manipulative, or unwanted activity on social networking platforms. Accurate identification of such accounts is challenging because bot and human accounts may share similar profile and behavioural characteristics. This work presents a machine-learning based framework for spambot prediction using account-level and behavioural features. Genuine-user and social-spambot datasets were merged and assigned Human and Bot labels. Missing values were replaced with zero, labels were encoded numerically, features were shuffled and normalized using Min-Max scaling, and the resulting dataset was divided into 80% training and 20% testing subsets. Several machine-learning classifiers, including Random Forest, SVM, Decision Tree, XGBoost, LightGBM, Logistic Regression, Extra Trees, Naive Bayes, and AdaBoost, were evaluated using accuracy, precision, recall, and F1-score. As an extension, Chi-Square SelectKBest feature selection was used to retain 20 features before training a CatBoost classifier. The extended CatBoost model achieved 99.567% accuracy, precision, recall, and F1-score on the test set. To improve model interpretability, LIME and SHAP were incorporated to examine the contribution of input features to predictions. The experimental results indicate that feature selection combined with CatBoost provides an effective approach for spambot prediction while explainability methods can provide additional insight into model decisions.

Key Words: Spambot detection, social networks, machine learning, CatBoost, Chi-Square feature selection, LIME, SHAP, explainable AI.

I. INTRODUCTION

Social networking platforms have become an important medium for communication, information sharing, and interaction among users. The large number of accounts and continuous user activity generated on these platforms also create opportunities for automated and malicious accounts, commonly referred to as social spambots. These accounts can imitate genuine users and generate unwanted or misleading activities, making the identification of bot accounts an important task for maintaining the reliability and security of social networking environments.

Traditional rule-based approaches for identifying spambots may not be sufficient because bot accounts can exhibit characteristics similar to genuine users. Machine learning techniques provide an effective alternative by learning patterns from account-level, behavioural, and content-related features. In this work, features such as followers, friends, favourites, verification information, profile attributes, hashtags, mentions, retweets, replies, URLs, vocabulary, sentence length, punctuation, polarity, and subjectivity are used for distinguishing genuine accounts from social spambots.

The proposed implementation combines genuine-user and social-spambot datasets to construct a binary classification dataset containing Human and Bot classes. The data is preprocessed by handling missing values, encoding the target labels, removing unnecessary attributes, shuffling the records, and applying Min-Max normalization. The resulting dataset is divided into 80% training data and 20% testing data, providing 5,544 training samples and 1,387 testing samples in the executed implementation. This preprocessing pipeline prepares the heterogeneous account and behavioural features for machine-learning classification.

To evaluate the effectiveness of different machine-learning approaches, the implementation compares several classifiers, including Random Forest, Support Vector Machine (SVM), Decision Tree, XGBoost, LightGBM, Logistic Regression, Extra Trees, Naive Bayes, and AdaBoost. Their performance is evaluated using accuracy, precision, recall, and F1-score. This comparative evaluation helps identify the most effective classification approach for the given spambot dataset.

As an extension to the baseline models, the proposed approach applies Chi-Square based SelectKBest feature selection to select the most relevant 20 features from the preprocessed feature set. These selected features are then provided to a CatBoost classifier for final classification. The purpose of this extension is to reduce the feature space while retaining informative characteristics that contribute to distinguishing bot and human accounts.

In addition to predictive performance, interpretability is considered an important part of the proposed framework. Therefore, LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (SHapley Additive exPlanations) are

incorporated to understand the contribution of individual features to model predictions. LIME provides local explanations for individual predictions, while SHAP provides feature contribution information that helps analyse the behaviour of the trained model.

The experimental implementation demonstrates that the proposed feature-selection and CatBoost based extension provides highly accurate spambot classification, achieving 99.567% accuracy, precision, recall, and F1-score on the evaluated test data. Thus, the proposed framework combines machine-learning based prediction with explainable AI to provide both effective and interpretable social-spambot detection.

II. RELATED WORK

A. Social Spambot Detection Using Machine Learning

Social-spambot detection has been widely studied using machine-learning techniques. Different studies have used account-level, content-based, and behavioural features to distinguish automated accounts from genuine users. Features such as follower count, friend count, favourites, verification status, profile characteristics, retweets, replies, hashtags, mentions, and URLs can provide useful information for identifying suspicious accounts. Classical machine-learning algorithms such as Decision Tree, Random Forest, Support Vector Machine, and Naive Bayes have been applied to these features for binary bot and human classification.

B. Ensemble Learning for Spambot Prediction

Ensemble learning methods have received considerable attention because they can capture complex relationships among social-network features. Random Forest combines multiple decision trees, while boosting methods such as AdaBoost, XGBoost, and LightGBM progressively improve classification performance by combining weak learners. These models are particularly suitable for structured tabular data containing heterogeneous account and behavioural features. In the present implementation, Random Forest, Decision Tree, XGBoost, LightGBM, AdaBoost, and Extra Trees are evaluated together with other baseline classifiers.

C. Feature Selection for Social-Bot Detection

Social-network datasets may contain a large number of features, and not all features contribute equally to classification. Irrelevant or less informative features can increase computational complexity and may affect model learning. Feature-selection techniques can therefore be used to retain the most informative attributes before classification. In this work, the Chi-Square SelectKBest technique is applied as an extension of the baseline classification pipeline. The implementation selects the 20 most relevant features, which are subsequently provided to the CatBoost classifier.

D. Explainable Artificial Intelligence for Spambot Detection

Although machine-learning models can achieve high classification performance, understanding the reasons behind their predictions is also important. Explainable Artificial Intelligence techniques provide information about the contribution of individual features to model decisions. LIME (Local Interpretable Model-Agnostic Explanations) can explain the prediction of an individual account, whereas SHAP (SHapley Additive exPlanations) provides feature-contribution information for model predictions. In this work, LIME and SHAP are incorporated to improve the interpretability of the spambot prediction system.

E. CatBoost-Based Classification

CatBoost is a gradient-boosting algorithm designed to provide effective classification performance while handling complex relationships in structured data. In the proposed implementation, CatBoost is used after Chi-Square feature selection rather than as an isolated baseline model. This combination forms the extended model, where the selected 20 features are used for training the CatBoost classifier. The performance of this extended approach is compared with the other machine-learning models evaluated in the implementation.

F. Summary of Existing Approaches

Existing approaches demonstrate that machine-learning models can effectively distinguish social spambots from genuine accounts using profile and behavioural features. However, predictive performance, feature reduction, model comparison, and interpretability are often considered separately. The present work combines these aspects in a single framework by comparing multiple classifiers, applying Chi-Square feature selection, using CatBoost as the extended classifier, and incorporating LIME and SHAP for explainability.

G. Research Gap

The existing studies on social-spambot detection have demonstrated the effectiveness of machine-learning techniques for distinguishing automated accounts from genuine users. Different studies have explored profile-based, behavioural, and content-related features, while several machine-learning and ensemble-learning algorithms have been applied for classification. However, the following requirements are still insufficiently addressed:

1. Comparative evaluation of multiple machine-learning classifiers under a common preprocessing and evaluation framework.

2. Selection of the most informative features from the available social-network feature set to reduce unnecessary or less relevant information before classification.
3. Integration of feature selection with an advanced boosting classifier for improving the effectiveness of spambot prediction.
4. Evaluation using multiple performance measures, including accuracy, precision, recall, and F1-score, rather than relying on a single performance metric.
5. Interpretability of spambot predictions, so that the important features contributing to the classification of an account as Human or Bot can be analysed.
6. Integration of prediction and explainability within a single framework that combines machine-learning classification with LIME and SHAP based analysis.

The proposed work addresses these requirements by developing an integrated spambot prediction framework. Multiple machine-learning classifiers are first evaluated using the same preprocessed dataset. Chi-Square SelectKBest is then applied to select the 20 most relevant features, followed by CatBoost classification as the extended approach. Finally, LIME and SHAP are incorporated to analyse and explain the model predictions. This framework therefore combines comparative machine-learning evaluation, feature selection, extended classification, and explainable AI within a single spambot detection system.

III. PROPOSED METHODOLOGY

A. Dataset Collection and Preparation

The implementation uses two types of data: genuine-user accounts and social-spambot accounts. The genuine-user data and its corresponding feature information are combined, and the social-spambot data and its feature information are similarly integrated. The two datasets are then merged to create a single dataset for binary classification. Each genuine account is assigned the label Human, whereas each social-spambot account is assigned the label Bot. The final merged dataset contains 6,931 records with the available account-level and behavioural features.

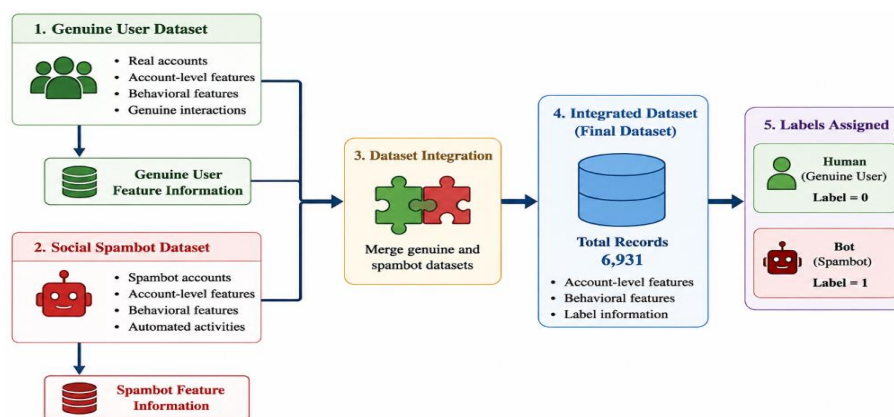
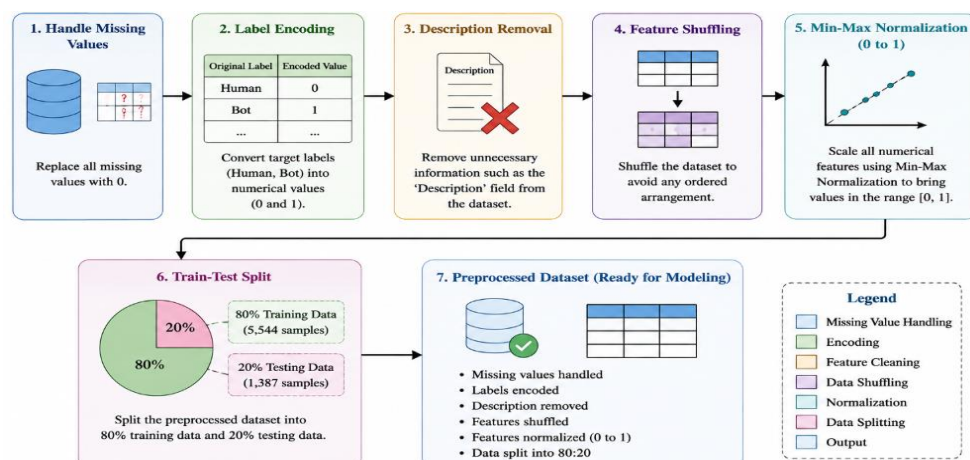


Fig. 1 – Dataset Collection and Integration

B. Data Pre-processing



The collected dataset is pre-processed before applying the classification algorithms. Missing values are handled by replacing them with zero. The target labels are converted into numerical form using Label Encoding. Unnecessary information, including the description field, is removed from the feature set before model training.

The feature data is shuffled to avoid an ordered arrangement of records. After pre-processing, Min-Max Normalization is applied so that the feature values are transformed into a common range between 0 and 1. This helps provide a consistent feature scale for the machine-learning models.

C. Machine-Learning Classification

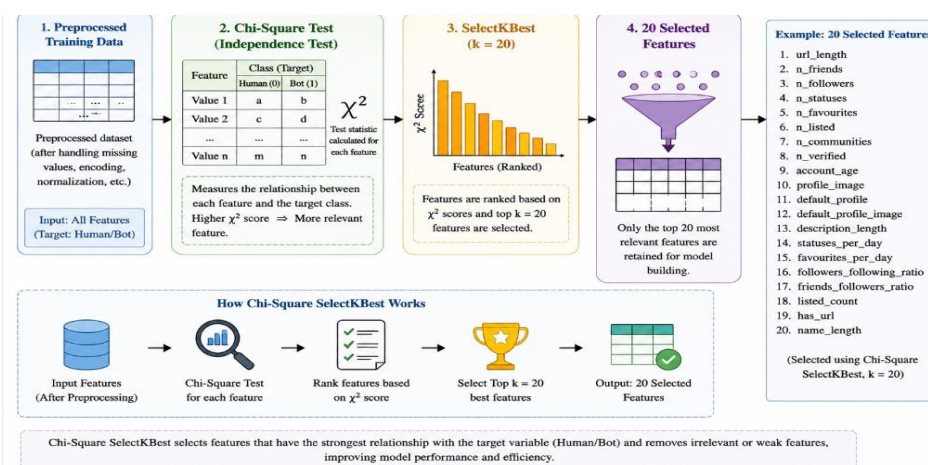
After preprocessing, several machine-learning algorithms are trained and evaluated for Human and Bot classification. The implemented algorithms include Random Forest, Support Vector Machine (SVM), Decision Tree, XGBoost, LightGBM, Logistic Regression, Extra Trees, Naive Bayes, and AdaBoost.

Each model is trained using the training dataset and tested using the testing dataset. The performance of the models is evaluated using accuracy, precision, recall, and F1-score. The comparative results are used to analyse the effectiveness of different machine-learning approaches for social-spambot prediction.

D. Feature Selection

To improve the extended classification process, Chi-Square based SelectKBest feature selection is applied to the preprocessed training data. The Chi-Square test evaluates the relationship between individual features and the target class and ranks the features according to their relevance.

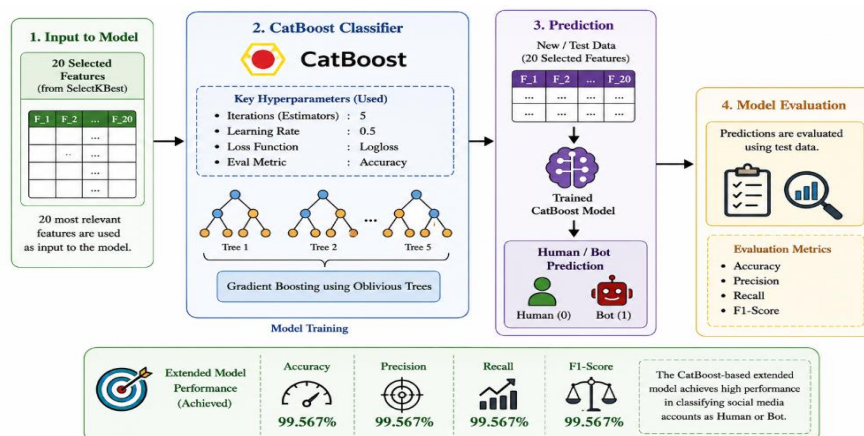
In the implementation, SelectKBest with $k = 20$ is used to select the 20 most relevant features. The feature-selection transformation learned from the training data is also applied to the test data.



E. CatBoost-Based Extended Model

The selected 20 features are used as input to the CatBoost classifier. The CatBoost model is trained using the selected training features and subsequently used to predict whether an account belongs to the Human or Bot class.

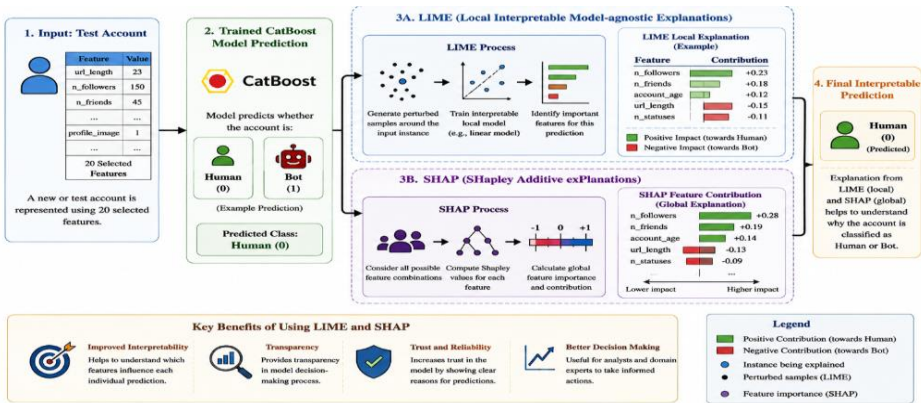
In the executed implementation, the CatBoost classifier uses 5 estimators and a learning rate of 0.5. The predictions are evaluated using accuracy, precision, recall, and F1-score. The extended CatBoost model achieves 99.567% accuracy, 99.567% precision, 99.567% recall, and 99.567% F1-score on the evaluated test data.



F. Explainable AI Using LIME and SHAP

To improve the interpretability of the prediction system, LIME and SHAP are incorporated into the proposed methodology. LIME is used to provide a local explanation for an individual account prediction by identifying the features that contribute to that prediction.

SHAP is used to analyse the contribution of features to the model predictions and provides a broader view of feature importance. The combination of LIME and SHAP helps to understand why an account is classified as Human or Bot.



IV.EXPERIMENTAL SETUP

A. Datasets and Preprocessing

The proposed spambot prediction framework is evaluated using a combined dataset consisting of genuine-user and social-spambot accounts. The genuine-user and social-spambot datasets are integrated to form the final classification dataset. The resulting dataset contains 6,931 records with account-level and behavioural features. Human accounts are assigned the Human class, while social-spambot accounts are assigned the Bot class.

Before model training, the data is preprocessed by converting the target labels into numerical values using LabelEncoder, replacing missing values with zero, removing the description field, and shuffling the feature records. Min-Max normalization is then applied to transform the feature values into the range [0, 1]. The processed dataset is divided into 80% training and 20% testing, resulting in 5,544 training samples and 1,387 testing samples.

B. Training Configuration

The training process includes multiple baseline machine-learning classifiers, namely Random Forest, Support Vector Machine, Decision Tree, XGBoost, LightGBM, Logistic Regression, Extra Trees, Gaussian Naive Bayes, and AdaBoost. Each classifier is trained using the training portion and evaluated using the independent testing portion.

For the proposed extension, Chi-Square SelectKBest with k = 20 is first applied to select the most relevant features. The selected features are then provided to a CatBoostClassifier configured with 5 estimators and a learning rate of 0.5. The trained model is evaluated using accuracy, precision, recall, and F1-score.

Parameter	Configuration
Dataset Records	6,931
Training Samples	5,544
Testing Samples	1,387
Train-Test Split	80%20%
Normalization	Min-Max Scaling [0, 1]
Baseline Models	9 ML Classifiers
Feature Selection	Chi-Square SelectKBest
Selected Features	20
Proposed Classifier	CatBoostClassifier
CatBoost Estimators	5
Learning Rate	0.5
Evaluation Metrics	Accuracy, Precision, Recall, F1-score
Explainability	LIME and SHAP

Table I - Experimental Configuration

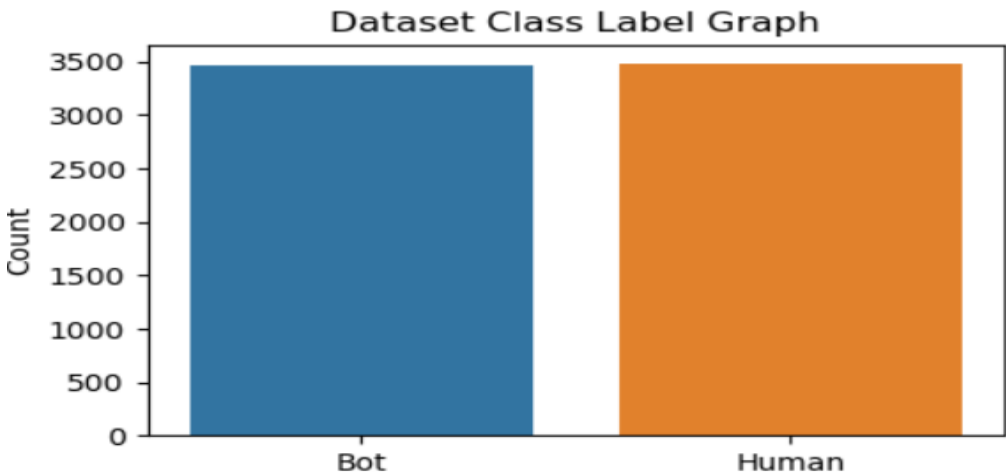
C. Explainability Configuration

The explainability stage is performed using LIME and SHAP. LIME generates a local explanation for an individual test instance and identifies the feature contributions associated with that prediction. SHAP uses a tree-based explainer to analyse feature contributions across the evaluated samples and generates a SHAP summary plot. These two techniques provide complementary explanations of the model's decision-making process.

V.EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

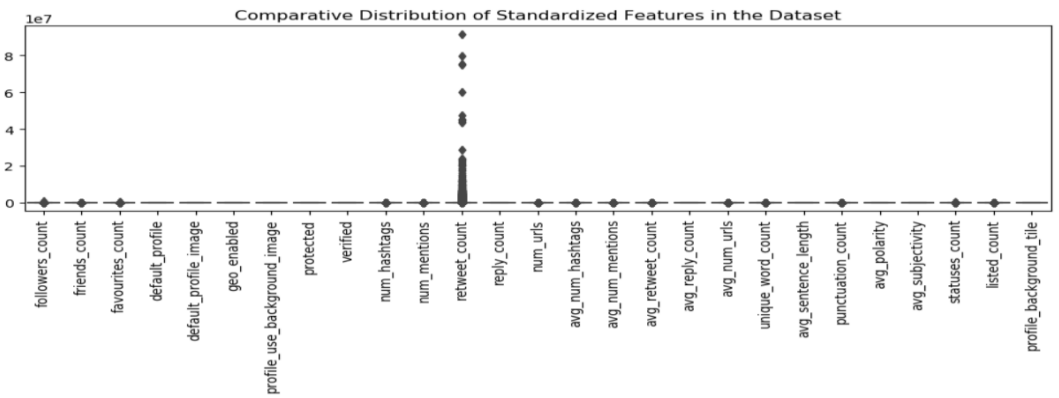
A. Dataset Class Distribution

The class distribution of the dataset was analysed before evaluating the classification models. Figure 1 presents the distribution of the two target classes, namely Bot and Human accounts. The graph shows that the number of Bot and Human accounts is approximately equal, indicating that the dataset has a nearly balanced class distribution. This balanced distribution provides a suitable basis for evaluating the performance of the classification models.



B. Feature Distribution Analysis

The distribution of the standardized features was analysed to understand the variation of the numerical attributes in the dataset. Figure 2 presents the comparative distribution of the standardized features using box plots. The visualization provides an overview of the feature distributions and highlights the presence of variations and extreme observations across different features. This analysis helps in understanding the characteristics of the feature set used for subsequent machine-learning classification.

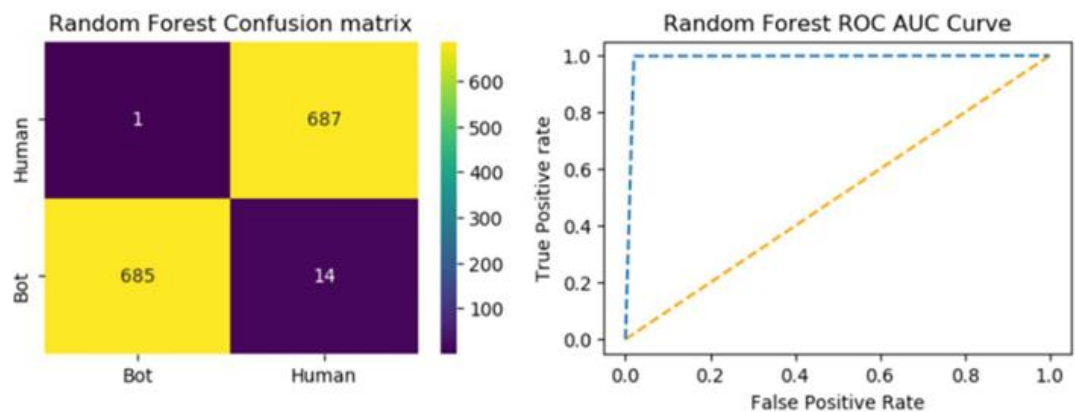


C. Feature Preprocessing and Training Feature Extraction

The dataset was preprocessed before applying the machine-learning models. The class labels were converted into numerical values using Label Encoding, and missing values were replaced with zero. The target label and description columns were removed from the feature set, and the remaining attributes were extracted as training features. The resulting feature matrix was used for subsequent machine-learning model training and evaluation.

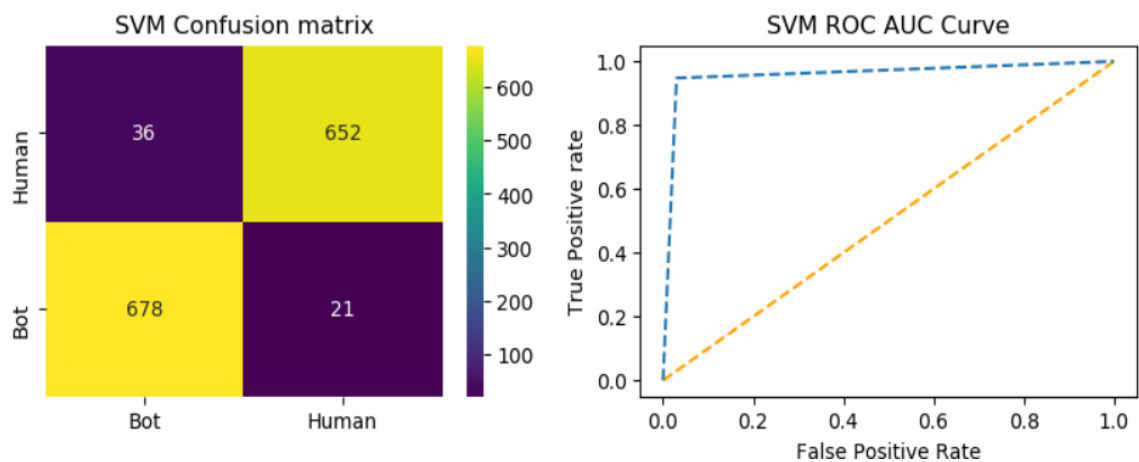
D. Random Forest Confusion Matrix

The Random Forest classifier achieved 98.919% accuracy, 98.929% precision, 98.926% recall, and 98.919% F1-score. The ROC curve shows strong classification performance, indicating that the model effectively distinguishes between human and bot traffic.



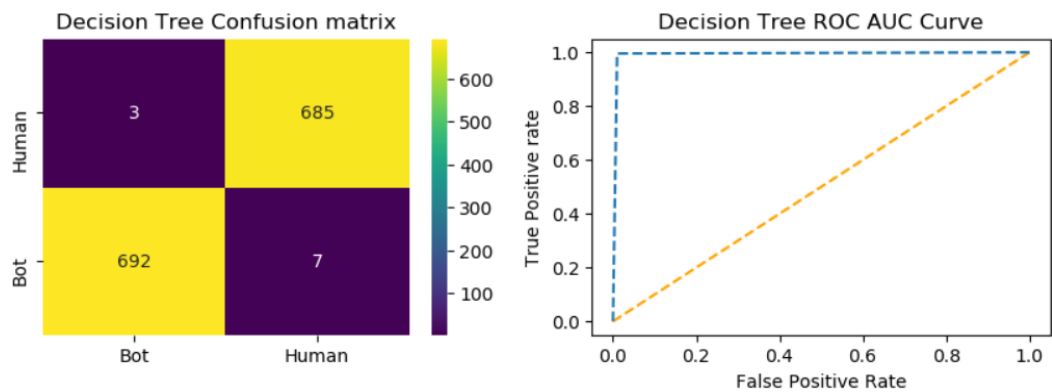
E. SVM Confusion Matrix

The SVM classifier achieved 95.89% accuracy, 95.919% precision, 95.882% recall, and 95.889% F1-score. The ROC curve indicates strong classification performance, showing that the SVM model can effectively distinguish between human and bot traffic.



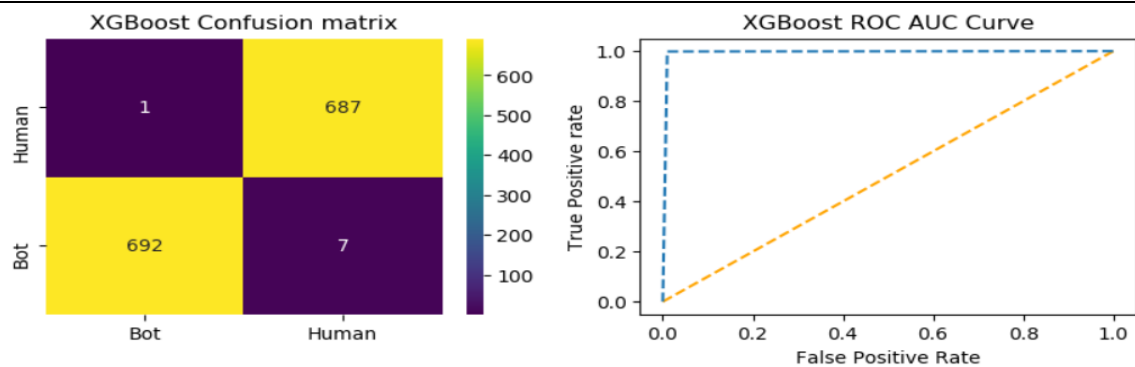
F. Decision Tree Performance Evaluation

The Decision Tree classifier achieved 99.279% accuracy, 99.278% precision, 99.281% recall, and 99.279% F1-score, demonstrating excellent classification performance in distinguishing human and bot traffic.



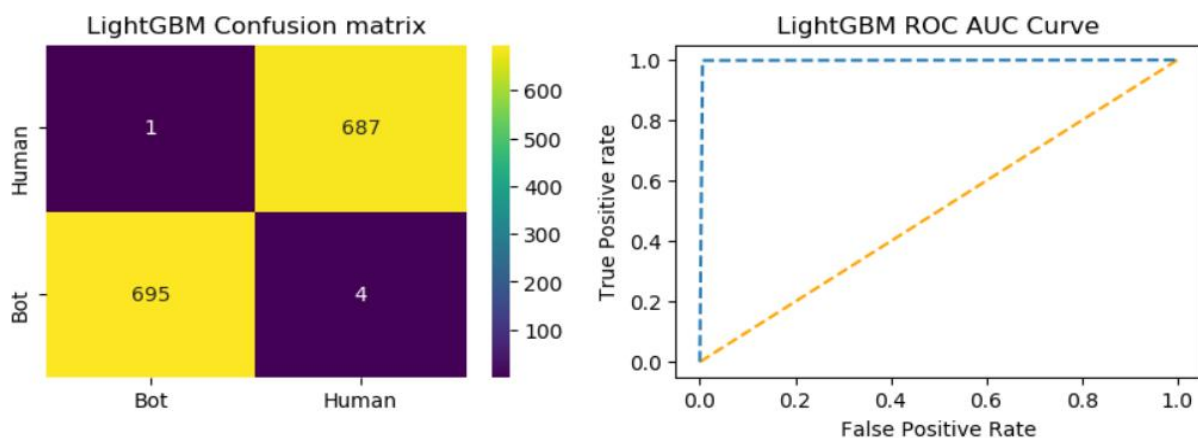
G. XGboost confusion matrices

XGBoost (Extreme Gradient Boosting) is an ensemble machine learning algorithm based on decision trees. It builds multiple trees sequentially, where each new tree focuses on correcting the errors of the previous trees. In this study, XGBoost was used for classification of the data into Human and Bot classes. The model achieved an accuracy of 99.423%, with 99.424% precision, 99.427% recall, and 99.423% F1-score. The confusion matrix and ROC curve also demonstrate the strong classification performance of the model.



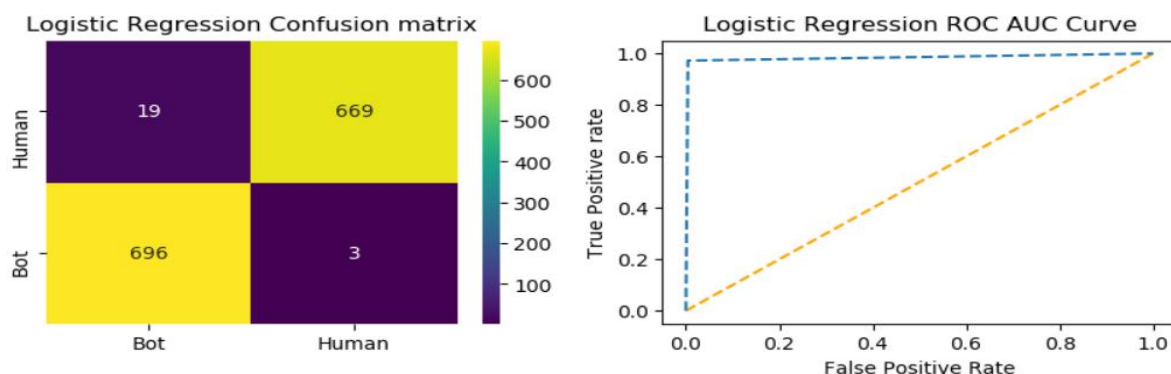
H. LightGBM Confusion matrices

LightGBM (Light Gradient Boosting Machine) is a gradient boosting algorithm designed for efficient and fast training on large datasets. It uses a tree-based learning approach and provides high classification performance with lower computational requirements. In this study, LightGBM was applied to classify Human and Bot samples. The model achieved an accuracy of 99.64%, with 99.639% precision, 99.641% recall, and 99.639% F1-score. The confusion matrix shows a small number of misclassified samples, while the ROC curve indicates excellent classification capability.



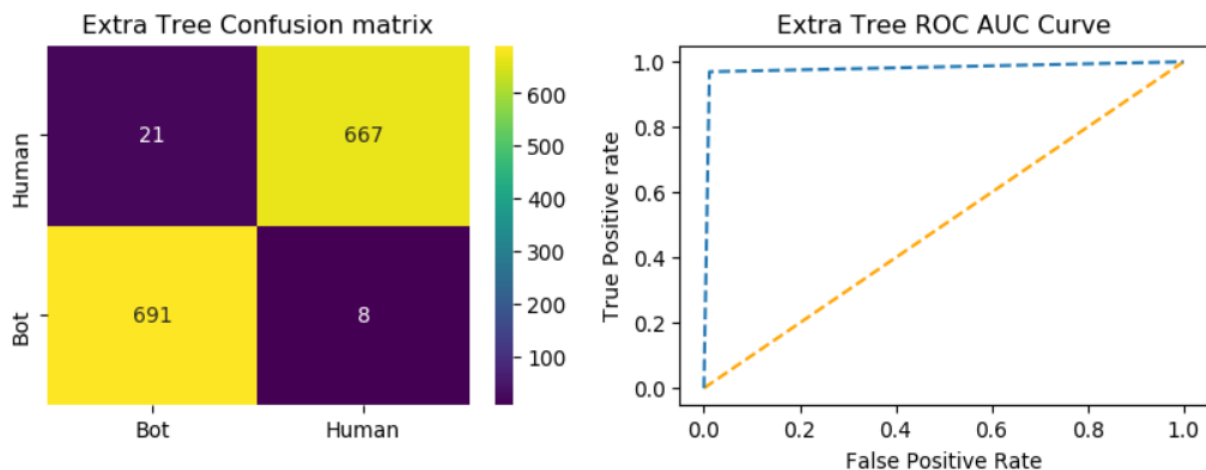
I. Logistic Regression Confusion matrix

Logistic Regression is a supervised machine learning algorithm used for binary classification. It estimates the probability of a sample belonging to a particular class and assigns the corresponding class label. In this study, Logistic Regression was used to classify Human and Bot samples. The model achieved an accuracy of 98.414%, with 98.448% precision, 98.405% recall, and 98.413% F1-score. The confusion matrix and ROC curve indicate good classification performance.



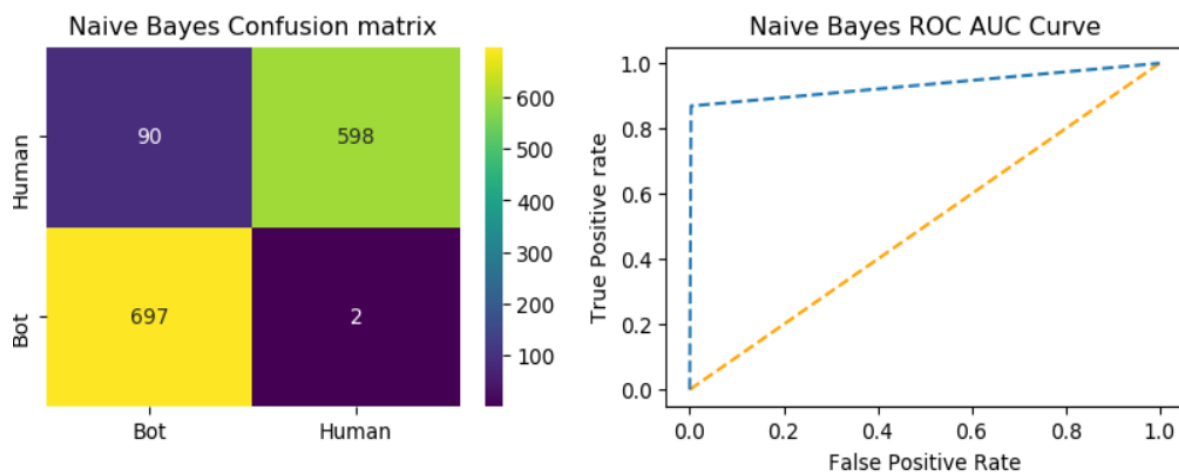
J. Extra Tree Confusion matrix

Extra Trees (Extremely Randomized Trees) is an ensemble learning algorithm that combines multiple randomized decision trees for classification. It introduces additional randomness while selecting features and split points, which helps in reducing overfitting and improving generalization. In this study, Extra Trees was used to classify Human and Bot samples. The model achieved an accuracy of 97.909%, with 97.933% precision, 97.902% recall, and 97.909% F1-score. The confusion matrix and ROC curve demonstrate the effectiveness of the model for classification.



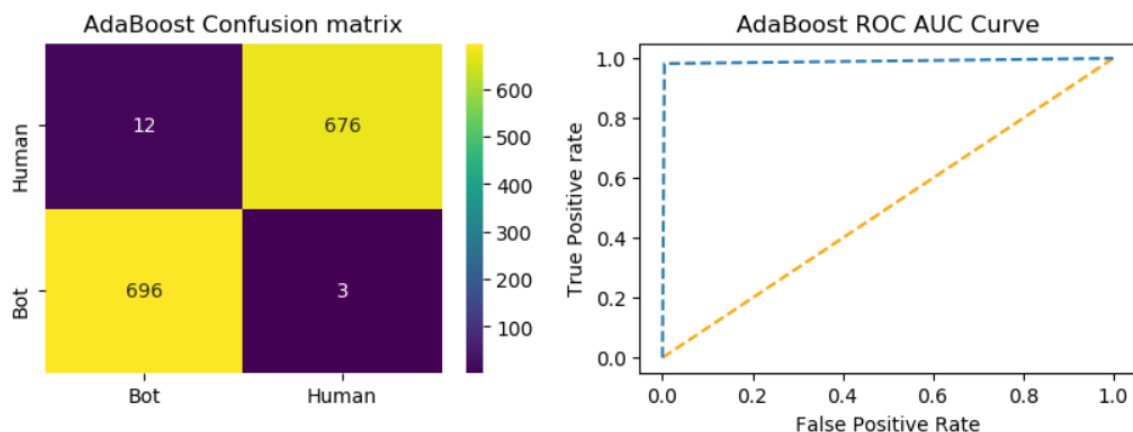
K. Naive Bayes Confusion matrix

Naive Bayes is a probabilistic supervised machine learning algorithm based on Bayes' theorem. It assumes that the features are conditionally independent of each other. In this study, Naive Bayes was used to classify Human and Bot samples. The model achieved an accuracy of 93.333%, with 94.115% precision, 93.316% recall, and 93.333% F1-score. The results indicate that the model provides effective classification performance.

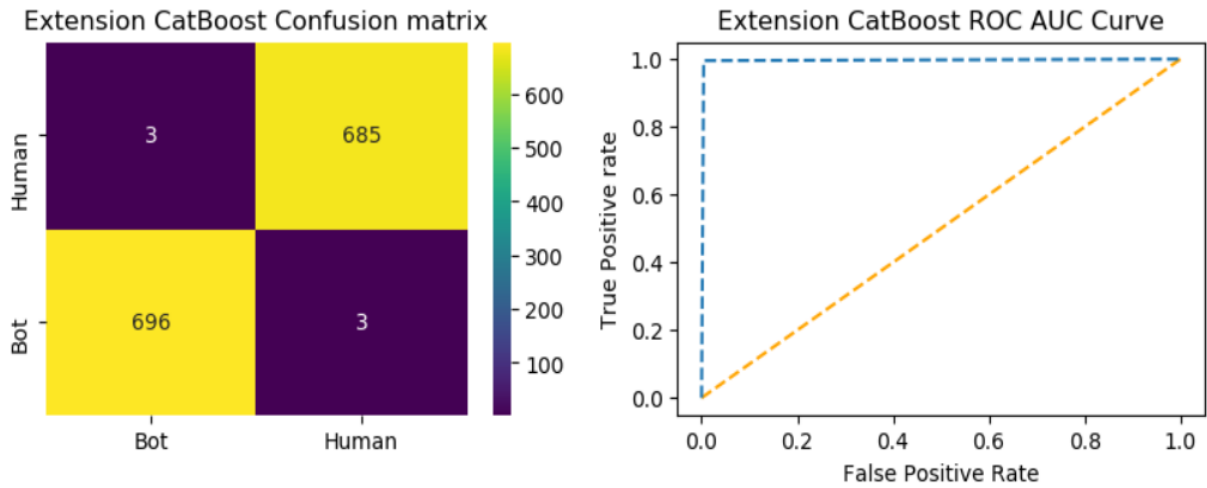


L. AdaBoost Confusion matrices

AdaBoost (Adaptive Boosting) is an ensemble learning algorithm that combines multiple weak learners to form a strong classifier. It gives more importance to incorrectly classified samples during successive training stages. In this study, AdaBoost was applied to classify Human and Bot samples. The model achieved an accuracy of 98.919%, with 98.932% precision, 98.913% recall, and 98.918% F1-score, demonstrating strong classification performance.



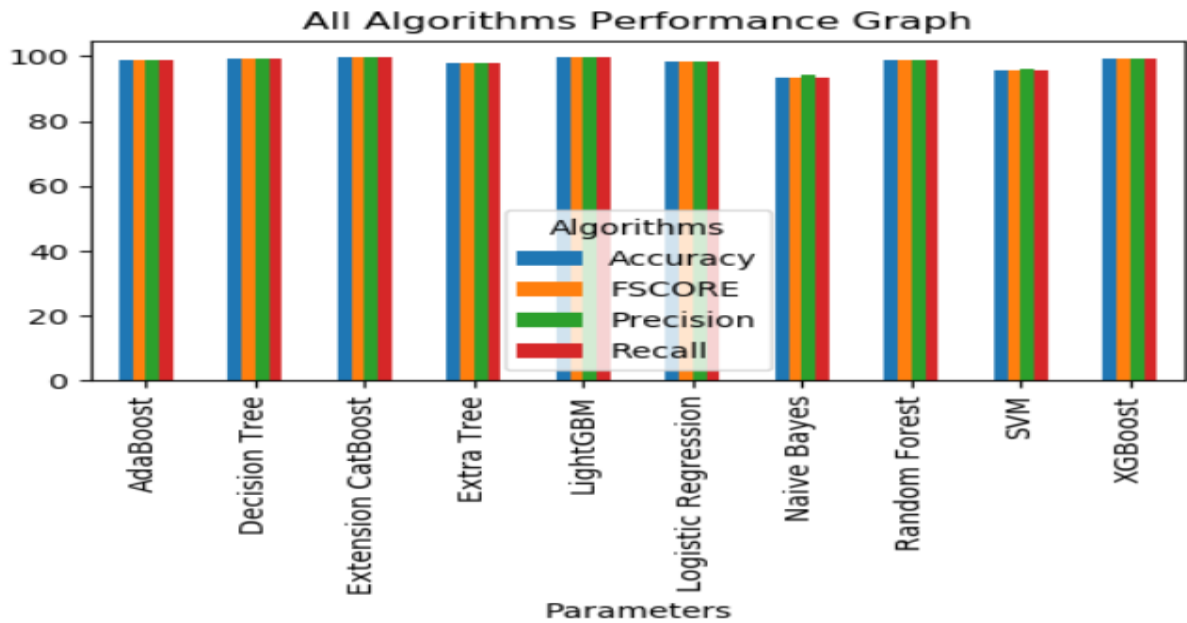
M. Extension CatBoost Confusion matrices



CatBoost (Categorical Boosting) is a gradient boosting algorithm based on decision trees. It is designed to provide efficient and accurate classification while handling categorical features effectively. In this study, CatBoost was used for Human and Bot classification. The model achieved an accuracy of 99.567%, with 99.567% precision, recall, and F1-score. The confusion matrix and ROC curve demonstrate the high predictive performance of the model.

N. Performance Comparison of machine learning Algorithm

Performance comparison of the different machine learning algorithms used for Human and Bot classification. The models were evaluated using accuracy, precision, recall, and F1-score. Among the evaluated models, LightGBM achieved the highest accuracy of 99.64%, followed by CatBoost (99.567%) and XGBoost (99.423%). Naive Bayes showed the lowest performance with an accuracy of 93.333%. Overall, the results demonstrate that the boosting-based models provide better classification performance compared with the other evaluated algorithms.



O. Overall Classification Performance

The performance of the proposed spambot detection framework was evaluated using multiple machine-learning classification algorithms. The models were trained using the preprocessed dataset and evaluated on the test dataset. The performance of each model was measured using four standard classification metrics: Accuracy, Precision, Recall, and F1-score. The comparative results obtained from the experiments are presented in Table II.

Table II. Performance Comparison of Machine-Learning Models

Table II shows that all the evaluated models achieved high classification performance. Among the baseline machine-learning models, LightGBM achieved the highest accuracy of 99.640%, followed by XGBoost with 99.423% and Decision

Tree with 99.279%. The Extension CatBoost model achieved an accuracy of 99.567%, demonstrating highly competitive performance for spambot detection.

P. Overall Discussion

The experimental results demonstrate that the evaluated machine learning algorithms are effective for distinguishing between Human and Bot samples. The models were evaluated using accuracy, precision, recall, and F1-score. The performance comparison shows that the boosting-based algorithms achieved better results compared with the other classification approaches.

Among the evaluated models, LightGBM achieved the highest accuracy of 99.64%, followed by CatBoost with 99.567% and XGBoost with 99.423%. Logistic Regression and AdaBoost also achieved high classification accuracy of 98.414% and 98.919%, respectively. Extra Trees achieved 97.909%, while Naive Bayes obtained 93.333% accuracy.

Overall, the results indicate that LightGBM, CatBoost, and XGBoost provide highly effective performance for Human and Bot classification.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	98.919	98.929	98.926	98.919
SVM	95.890	95.919	95.882	95.889
Decision Tree	99.279	99.278	99.281	99.279
XGBoost	99.423	99.424	99.427	99.423
LightGBM	99.640	99.639	99.641	99.639
Logistic Regression	98.414	98.448	98.405	98.413
Extra Trees	97.909	97.933	97.902	97.909
Naive Bayes	93.367	94.115	93.316	93.333
AdaBoost	98.919	98.932	98.913	98.918
Extension Cat Boost	99.567	99.567	99.567	99.567

VI. CONCLUSION AND FUTURE WORK

A. Conclusion

This paper presented a machine learning-based approach for the classification of Human and Bot samples. Multiple supervised machine learning algorithms were implemented and evaluated to identify the most effective model for distinguishing between human-generated and bot-generated samples. The evaluated algorithms include Logistic Regression, Extra Trees, Naive Bayes, AdaBoost, XGBoost, LightGBM, and CatBoost, along with other classification models used in the experimental analysis.

The experimental results demonstrate that the evaluated models achieved high classification performance. Among the tested algorithms, LightGBM achieved the highest accuracy of 99.64%, followed by CatBoost with 99.567% and XGBoost with 99.423%. Logistic Regression and AdaBoost achieved accuracies of 98.414% and 98.919%, respectively, while Extra Trees and Naive Bayes achieved 97.909% and 93.333%. The precision, recall, and F1-score results further confirm the effectiveness of the models in Human–Bot classification.

The confusion matrix and ROC-AUC analysis also showed that the high-performing models produced a small number of misclassified samples and provided strong class discrimination. The overall performance comparison indicates that boosting-based ensemble models, particularly LightGBM, CatBoost, and XGBoost, are more effective for the given classification task.

Overall, the findings demonstrate that machine learning techniques can provide an accurate and reliable solution for Human–Bot classification. The proposed experimental framework can be further enhanced in future work by using larger and more diverse datasets, advanced feature selection techniques, hyperparameter optimization, and additional ensemble learning methods to improve the robustness and generalization of the classification system in real-world applications.

B. Future work

Although the proposed machine learning framework demonstrates promising performance, several opportunities remain for future research. First, advanced deep learning techniques, including convolutional neural networks, recurrent neural networks, and transformer-based architectures, can be explored and compared with the machine learning models

evaluated in this study. Such approaches may further improve prediction performance when sufficient training data are available.

Second, future studies can focus on systematic hyperparameter optimization using techniques such as Grid Search, Random Search, or Bayesian Optimization to identify optimal model configurations. Advanced ensemble learning methods, such as stacking and weighted voting, can also be investigated to improve the robustness and generalization capability of the predictive system.

Furthermore, additional feature engineering and feature selection techniques can be explored to identify the most informative features and reduce model complexity. Explainable Artificial Intelligence (XAI) techniques such as SHAP and LIME can also be incorporated to improve the interpretability and transparency of model predictions.

Future research should additionally evaluate the proposed framework on independent and diverse datasets to assess its generalizability and robustness in real-world scenarios. Finally, the best-performing model can be integrated into a real-time application or API, enabling practical deployment and continuous evaluation in real-world environments.

References

1. S. N. Tirumala Rao, Sireesha Moturi, Najeer Shaik, Suneetha Mothe, Sistla V. S. Manikanta Rohit, Ravi Lakshmi Sri Harsha, and Dodda Venkata Reddy, "Fake Profile Detection Using Machine Learning,"2025.
2. Rajat Soni, Shubham Jain, and Ankit Kumar, "Fake Twitter Followers Detection Using Machine Learning Approach,"2020.
3. M. Santhoshi, S. Sailaja, and J. Jyotsna, "Deep Learning Approach for Identification of Fake Profiles in Social Media,"2023
4. Murtala Aminu Buhari, Salau Ayodeji Olalekan, Azadeh Mohammadi, and Mo Sarace, "A Comparative Study on the Characteristics and Behaviours of Social Media Users in Response to Fake News on X (Twitter),"2024.
5. Sarith Imaduwege, P. P. N. V. Kumara, and W. J. Samaraweera, "Importance of User Representation in Propagation Network-based Fake News Detection: A Critical Review and Potential Improvements,"2022.
6. Andrea Cresci, Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi, and Maurizio Tesconi, "Detecting Fake Profiles on Social Networks: A Systematic Investigation," 202
7. S. K. Lakshmanaprabu, K. Shankar, M. Ilayaraja, S. S. Kumar, and N. Arunkumar, "Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques,"2019.
8. Matsumoto Hayato, Soh Yoshida, and Mitsuji Muneyasu, "Flexible Framework to Provide Explainability for Fake News Detection Methods on Social Media,"in 2022.
9. A. S. Bandi, P. S. Reddy, and K. R. Reddy, "Detection of Fake Users in Twitter Using Network Representation and NLP,"2022.
10. Sara Asghari, Mostafa Haghiri Chehreghani, and Morteza Haghiri Chehreghani, "On Using Node Indices and Their Correlations for Fake Account Detection,"in 2022
11. Ryo Onuma, Hiroaki Kaminaga, Hiroki Nakayama, Youzou Miyadera, Keito Suzuki, and Shoichi Nakamura, "Analysis of Articles that Correct Other Posts on Social Media Aimed at Promoting the Experience in Examining Fakes,"in 2022.
12. Kai Shu, Suhang Wang, and Huan Liu, "Nothing Stands Alone: Relational Fake News Detection with Hypergraph Neural Networks," 2020.