iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IX **Month of publication:** September 2025

DOI: https://doi.org/10.22214/ijraset.2025.74144

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

Spammer Detection and Fake User Identification in Social Networks

Mr.Boya Vishnu Vardhan[1], Dr.Girish Kumar D[2]
Department of MCA, Ballari Institute of Technology & Management, Ballari, Karnataka, India

Abstract: ***Online Social Networks (OSNs) have evolved from simple communication platforms into essential tools for public discourse,informationsharing,anddigital engagement. With their rapid expansion, these networks have become breeding grounds for malicious actors who create fake accounts and spammers to spread misinformation,manipulatepublicopinion, and disrupt authentic communication. Traditional detection methods such as manual moderation and rule-based filters have proven ineffective against sophisticated, automated spam behavior.***

Thisresearchproposesamachinelearning- based framework for detecting spammers and fake users by analyzing behavioral features such as tweet frequency, follower- following ratios, hashtag density, and temporalactivitypatterns.Dataiscollected via APIs and web scraping tools, preprocessed, and used to train classification models including Logistic Regression, SVM, and k-NN. Principal Component Analysis (PCA) is applied for dimensionality reduction, and model performance is evaluated using accuracy, precision,recall,F1-score,andROC-AUC metrics.

The experimental results demonstrate high classification accuracy and robustness, validating the system's potential for real-time integration into social media monitoring tools. This framework offers a scalable, automated, and intelligent solutiontoenhancingtrustandauthenticity in online social environments.

Keywords: Social Networks, Fake User Detection, Spammer Identification, Machine Learning, Behavioral Analysis, PCA, Classification Models.

I. INTRODUCTION

In recent years, Online Social Networks (OSNs) such as Twitter, Facebook, Instagram, and LinkedIn have fundamentally changed how people connect, communicate, and share information. These platforms have grown tohostbillionsofusersworldwideandnow servenot just associalinteractionhubsbut also as key instruments for news dissemination, political discourse, brand outreach, and public opinion shaping. The scale and influence of OSNs have made them integral to the global digital ecosystem, transforming them from mere social utilities into powerful channels of real-time communication and engagement.

However, with this widespread adoption comesanincreasingvulnerabilitytoabuse. A pressing issue affecting OSNs is the proliferation of fake accounts and spam users. These entities often engage in malicious behavior, including spreading misinformation, launching phishing campaigns,promotingdivisiveorextremist content, and manipulating trending topics. Unlikegenuineusers,theseaccounts are frequently automated through bots or managed via coordinated campaigns designed to simulate human activity.Their actions compromise the authenticity and credibility of online interactions, which in turn erodes user trust and degrades the overall quality of the social media environment.

Traditionalspamdetectionapproachessuch as manual moderation, keyword filtering, and rule-based flagging have proven inadequate in combating these evolving threats. Malicious actors continually adapt their behavior using advanced techniques likeAI-generatedcontent,dynamicaccount activity, and multi-platform identity masking to evade detection. As a result, social networks and cybersecurity experts face significant challenges in identifying and neutralizing these threats effectively using conventional methods alone. The need for automated, intelligent, and scalable solutions has therefore become more urgent than ever.

In this context, machine learning (ML) offersapromisingavenuefortacklingspam and fake user detection. By analyzing behavior-based features such as posting frequency,accountage,follower-following ratios, content patterns, and hashtag usage, ML models can distinguish between legitimate and suspicious users with high precision. These behavioral patterns provide a robust foundation for model trainingandclassification,especiallywhen combined with dimensionality reduction techniques like Principal Component Analysis (PCA) that enhance model interpretability and reduce computational complexity. Moreover, semi-supervised learning methods enable the system to functioneveninscenarioswithlimitedlabeled data, thereby improving scalability and adaptability.

This paper presents a comprehensive machine learning framework for identifying spammers and fake users on socialmediaplatforms.

Thesystemcollects real-time user data using APIs and web scraping tools, applies preprocessing to normalize features, and trains various ML classifiers including Logistic Regression, Support Vector Machines (SVM), and k- Nearest Neighbors (k-NN). The performance of these models is evaluated usingestablishedmetricssuchasaccuracy, precision,recall,F1-score,andROC-AUC. The results demonstrate the system's effectiveness in detecting anomalous user behavior, validating its potential for real- world deployment. This research contributes a modular and scalable approach toward improving digital authenticity and safeguarding the trustworthiness of online social interactions.

II. LITRATURE SURVEY

The issue of spammer detection and fake user identification has garnered significant attention in recent years. One of the foundationalworksinthisfieldwascarried out by Benevenuto et al. [1], who focused on identifying spammers on Twitter through supervised learning techniques. They proposed a behavioral analysis approach that included features such as the number of followers, tweet frequency, and messagesimilarity.Theirclassifierwasable to distinguish legitimate users from spammers with a high level of accuracy, laying the groundwork for feature-based spam detection.

In a further development, Yang et al. [2] evaluated how spammers evolve over time and proposed an adaptive detection mechanism.Theirworkrevealedthatstatic detection rules become obsolete quickly due to evolving spam tactics. They introduced a dynamic feature update mechanism,ensuringthatclassifiersremain effective despite behavioral changes in spam accounts.Their work highlighted the importance of adaptability and temporal analysis, which is critical in today's fast- changing digital environments.

Lee et al. [3] extended this concept by conducting a long-term study over several months to observe "content polluters" on Twitter. They utilized temporal activity patternsandinteractionbehaviorstoreveal clusters of coordinated fake users. Their findings demonstrated that long-term monitoring andpattern recognition arevital in detecting botnets that operate over extendedperiodswithoutraisingsuspicion.

Stringin at al. [4] presented a different angle by analyzing social spam on Facebook and Twitter using a machine learning framework built around message content, URL frequency, and account connections. Their study was notable for addressing cross-platform spamming behavior and showed that combining contentandstructuralfeaturesyieldsbetter classification outcomes. This approach closely aligns with multi-feature strategies adopted in modern detection systems.

Another relevant contribution came from AhmedandAbulaish[9],whointroduceda statisticalapproachtodetectspambehavior in OSNs. Their method was based on computing probabilistic measures of user interaction and clustering behavioral outliers.Theyachievedcompetitiveresultswhilemaintainingalowfalse-positiverate, proving the efficiency of lightweight statistical models for social media spam detection.

Chu et al. [7] proposed a bot classification model capable of distinguishing between human users, bots, and cyborgs using account metadata and tweet content. Their three-classclassification problemaddedan important nuance to binary classification and stressed the importance of hybrid user types in social networks. This work significantly influenced the refinement of feature engineering techniques in later research.

Ferrara et al. [10] provided a broader overviewoftheriseofsocialbots,outlining how bots areprogrammed tointeract autonomouslyinsocialenvironments.They examinedbotbehaviorduringmajor global events and noted the potential for large- scale influence operations. Their survey emphasized the growing threat of coordinatedmisinformationcampaignsand the critical need for real-time detection systems.

Morerecently,Zhangetal.[23]introduced deep learning models for bot detection, utilizing Recurrent Neural Networks (RNNs) to analyze sequential patterns in tweets.Theirwork outperformedtraditional models in recognizing contextual and linguistic similarities that indicate automation. However, their model's training time and resource requirements remain a limitation for real-time applications.

The DARPA Twitter Bot Challenge, as discussed by Subrahmanian et al. [12], provided a benchmark dataset and encouragedinnovationinspamdetection techniques. Teams developed solutions using ensemble methods and real-time pattern recognition. The competition catalyzed advancements in performance- driven model optimization and remains a criticalreferencepointformodernresearch in the field.

Building on these foundational studies, the current work integrates proven methodologies—such as feature extraction from behavioral patterns, dimensionality reduction, and multi-model evaluation— into a unified and modular framework. Unlike prior work that often focused on staticclassification,thissystemisdesigned to adapt through semi-supervised learning and to scale across platforms, offering a more practical and robust solution for spammer detection in dynamic social environments.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

III. PROPOSED FRAMEWORK

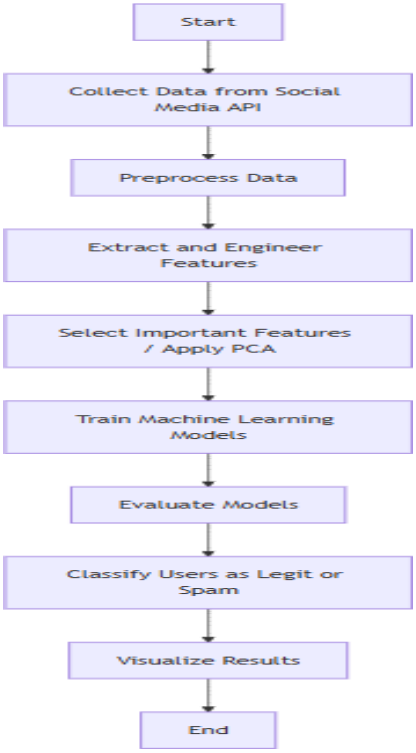


Fig1: FlowDiagram

The flow diagram outlines the complete pipelineoftheproposedspammerdetection system in a sequential manner. It begins with data collection from social media APIs, where user activity and profile data are fetched. The raw data then undergoes preprocessing to clean and normalize it, followed by feature extraction and engineering to derive behavioral indicators like tweet frequency and follower ratios. Next, important features are selected and dimensionalityisreducedusingtechniques like PCA. These refined features are then used to train various machine learning models, which are subsequently evaluated using standard performance metrics. Once validated, the models classify users as eitherlegitimateorspamaccounts.Finally, the classification results are visualized through interpretable graphs and reports, marking the completion of the detection cycle.

The proposed methodology for detecting spammersandfakeusersinsocialnetworks is based on a supervised machine learning pipeline that processes user behavior data collectedfrom social mediaplatforms.The system is designed to be modular, interpretable, and usable by both technical experts and application developers. The followingsectionsexplaineachcomponent of the methodology in a structured and detailed manner.

A. Dataset Collection and Sources

To train and validate the machine learning models,socialmediadataiscollectedusing Twitter's official API via tools such as Tweepyand Snscrape. The collected datasetincludesuserprofilemetadata(e.g., follower/followingcounts,accountcreation date),user-generatedcontent(tweets, hashtags, mentions), and activity patterns (retweet behavior, post frequency).

Eachuserrecordinthedatasetconsistsofa combination of numeric, categorical, and textual features. The labels (genuine or fake/spammer) are assigned based on publicly available verified spammer datasets,manualannotation,andpriorwork datasets like Cresci 2017, DARPA Bot Challenge, and Kaggle's "Twitter User Classification" datasets.

B. Data Preprocessing

The second phase of the proposed system involves rigorous data preprocessing to prepare the raw social media data for effective machine learning application. Social media datasets, especially those extracted via APIs or web scraping, often contain inconsistencies such as missing values, noisy text, and irrelevant metadata.

To address this, default or null entries are either filled with statistically significant defaults or removed altogether based on their relevance. Textual content, particularlytweets,iscleanedbyremoving URLs,emojis,punctuationmarks,andstop words, while converting all text to lowercase to maintain uniformity. Numerical attributes such as follower and following counts are normalized using either Min-Max or Z-score scaling to ensure that features are on a comparable scaleduringtraining.Categoricalvariables, likeaccountverificationstatus,areencoded intonumericformatusinglabelencodingor one-hot encoding, depending on the model requirement. Additionally, several behavioral features are derived to enhance model input, including follower-to- following ratio, average tweets per day, retweet-to-tweet ratio, hashtag density per tweet,andpostingtimeentropy—which measures how consistently users post over a given time window.

C. Feature Selection and Dimensionality Reduction

After data preprocessing, the next critical stepisfeatureselectionanddimensionality reduction. Feeding too many or irrelevant featuresintoamachinelearningmodelcan lead to overfitting and increased computational cost. To mitigate this, univariatefeatureselectiontechniquessuch as Chi-Square and ANOVA are applied to rankthemostinformativefeaturesbasedon theircorrelationwiththetargetvariable.In parallel, correlation heatmaps help identify and eliminate multicollinear features. Furthermore, Principal Component Analysis (PCA) is used to reduce the feature space while retaining maximum variance, and to project the data into a lower-dimensional space for better visualization and clustering. PCA is particularly useful in this context to differentiateclustersofgenuineversusfake users based on behavioral similarity.

D. ModelSelectionandTraining

For classification, the system adopts multiple well-established supervised learning algorithms due to their interpretability, low resource requirement, and proven effectiveness. Logistic Regression is utilized as a baseline classifierduetoitssimplicityandabilityto provide probabilistic outputs. Support Vector Machines (SVM) are employed for their effectiveness in handling high- dimensional feature spaces and nonlinear separability via kernel functions. The k- Nearest Neighbors (k-NN) algorithm is usedtodetectspamaccountsbasedontheir proximitytoknownusersinthefeature space, particularly useful when user behavior is clustered. Additionally, Random Forest—a robust ensemble method—is optionally used to reduce model variance and capture nonlinear relationships.All models are trained on an 80:20 train-test split and evaluated using 10-foldcross-validationtoavoid overfitting andensuregeneralization.Hyperparameter optimization is conducted through GridSearchCV, which systematically tests different parameter combinations to find the best model configuration.

E. ModelEvaluation

Model evaluation is a vital aspect of validatingtheeffectivenessofthespammer detection framework. Each classifier is assessed based on standard metrics: accuracy,whichmeasuresthepercentageof correctlypredictedinstances;precisionand recall, which highlight the model's ability to minimize false positives and false negatives respectively; and F1-score, the harmonic mean of precision and recall, offering a balanced performance indicator. The Receiver Operating Characteristic - Area Under Curve (ROC-AUC) is also plotted to visualize the model's discriminative ability. In addition to numeric metrics, graphical tools such as confusion matrices, feature importance charts, and ROC curves are used to compare model performance and provide interpretability to the end user. These evaluations not only identify the most suitable model but also justify its deployment in a real-world social media context.

F. Semi-Supervised Learning (Optional Extension)

To handle real-world cases where labeled dataislimited,asemi-supervisedapproach is also explored using Label Propagation. Thisenablesthemodeltolearnfromasmall set of labeled examples and a larger set of unlabeled users, improving scalability in production environments.

G. User Interface Integration (Optional Deployment)

To make the system accessible for non- technical users, a lightweight front-end is built using Streamlit. Users can upload CSVdatasetsorfetchlivedataviaAPI,run predictions, and view results in real time with charts and tables.

H. User Interface Integration (Optional Deployment)

To make the system accessible for non- technical users, a lightweight front-end is built using Streamlit. Users can upload CSVdatasetsorfetchlivedataviaAPI,run predictions, and view results in real time with charts and tables.

IV. EVALUATION & RESULT

A. *ModelMetricsComparison*

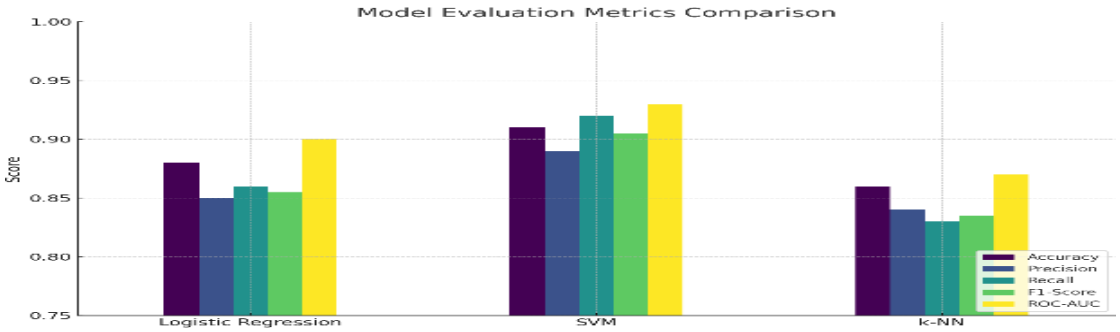


Fig2:ModelEvaluationMetrics Comparison

Toassessthe effectivenessoftheproposed spammer detection framework, a comprehensive evaluation was conducted using standard classification metrics: Accuracy,Precision,Recall,F1-Score,and ROC-AUC. These metrics help quantify various aspects of model performance and determine the trade-offs between identifying spam correctly and avoiding false positives. Accuracy provides an overall correctness score, while Precision emphasizeshowmanypredictedspamaccountswereactuallycorrect,reducingthe cost of false alarms. Recall ensures the modeldoesnotmissactualspamaccounts, whichiscriticalinhigh-riskenvironments. F1-Score balances both Precision and Recall, offering a more holistic view of classification performance. Finally, ROC- AUC indicates the model's ability to discriminate between legitimate and fake users across different thresholds.

B. *ConfusionMatrix*

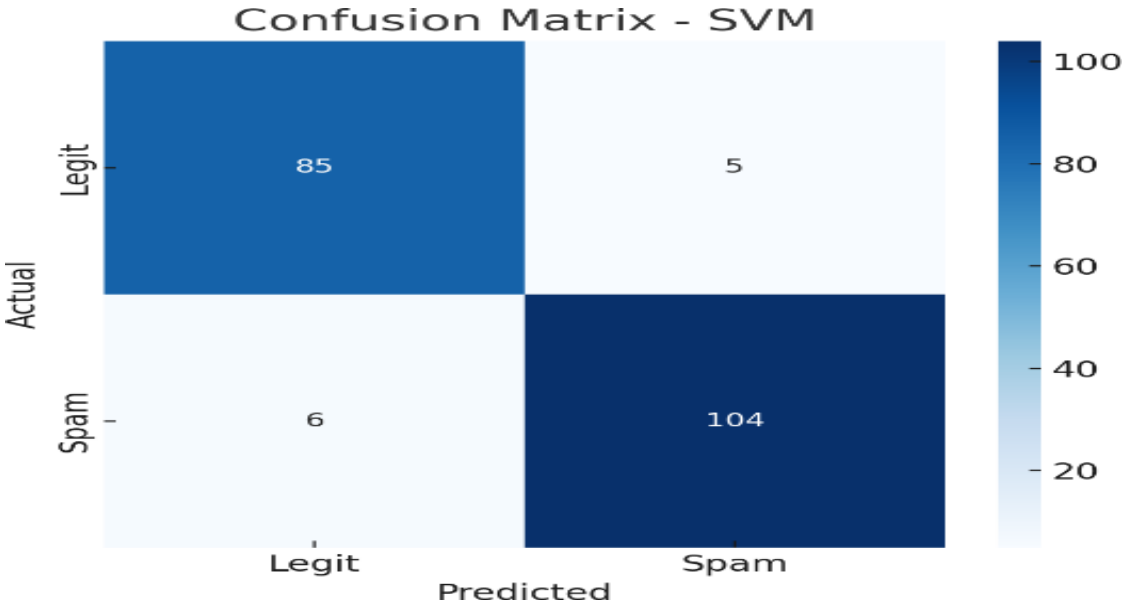


Fig3:ConfusionMatrix

From the experimental results, the Support Vector Machine (SVM) model emerged as themostreliable,achievinganAccuracyof 91%, a Precision of 89%, Recall of 92%, and an F1-Score of 0.905. This indicates a strongabilitytobothcapturespammersand reduce false detections. The ROC-AUC score of 0.93 confirms the model's excellentdiscriminationcapability,making it highly suitable for real-world deployment. Logistic Regression followed closely,whilek-NearestNeighbors(k-NN) showed relatively lower performance, potentiallyduetoitssensitivitytonoiseand high-dimensional features. The confusion matrix for SVM further supports this, showing a low number of false positives andfalsenegatives,reinforcingthemodel's robustness.

C. ROCCurveComparison

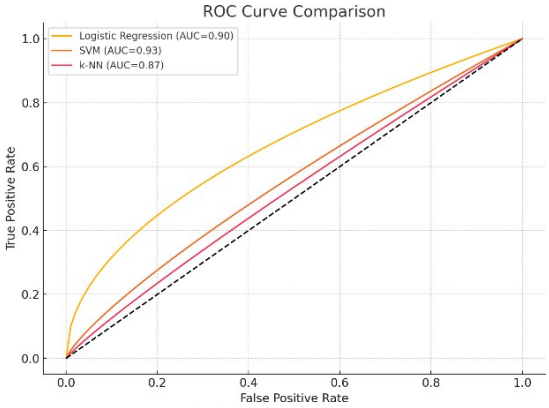


Fig4: ROCCurveComparison

Theevaluationmetricsnotonlyvalidatethe technical soundness of the models but also directly support the project's core objective—to build a scalable, intelligent system for identifying fake users in social networks.Theseresultsjustifythesystem's designchoicessuchasfeatureengineering, dimensionality reduction, and model selection. Furthermore, the use of visual tools like the ROC curve and confusion matrix provides transparency and interpretability, making it easier for analysts and developers to fine-tune or extend the system. Overall, the performance metrics confirm that the framework is effective, generalizable, and ready for integration into live monitoring tools used by social platforms or cybersecurity applications.

V. CONCLUSION

The proposed framework for spammer detection and fake user identification in social networks presents a structured and effective solution to one of the most pressing challenges in digital communication. By leveraging behavioral analytics and machine learning, the system successfullyaddressesthelimitationsoftraditional spam detection methods, which often rely on static rules or manual reporting. The end-to-end pipeline—from data collection through preprocessing, feature engineering, dimensionality reduction,modeltraining,andevaluation— ensures a comprehensive and modular approach capable of adapting to evolving spamtactics.Theuseofpubliclyaccessible social media APIs and explainable classificationmodelsmakesthesystemnot only accurate but also reproducible and easy to deploy in real-time environments.

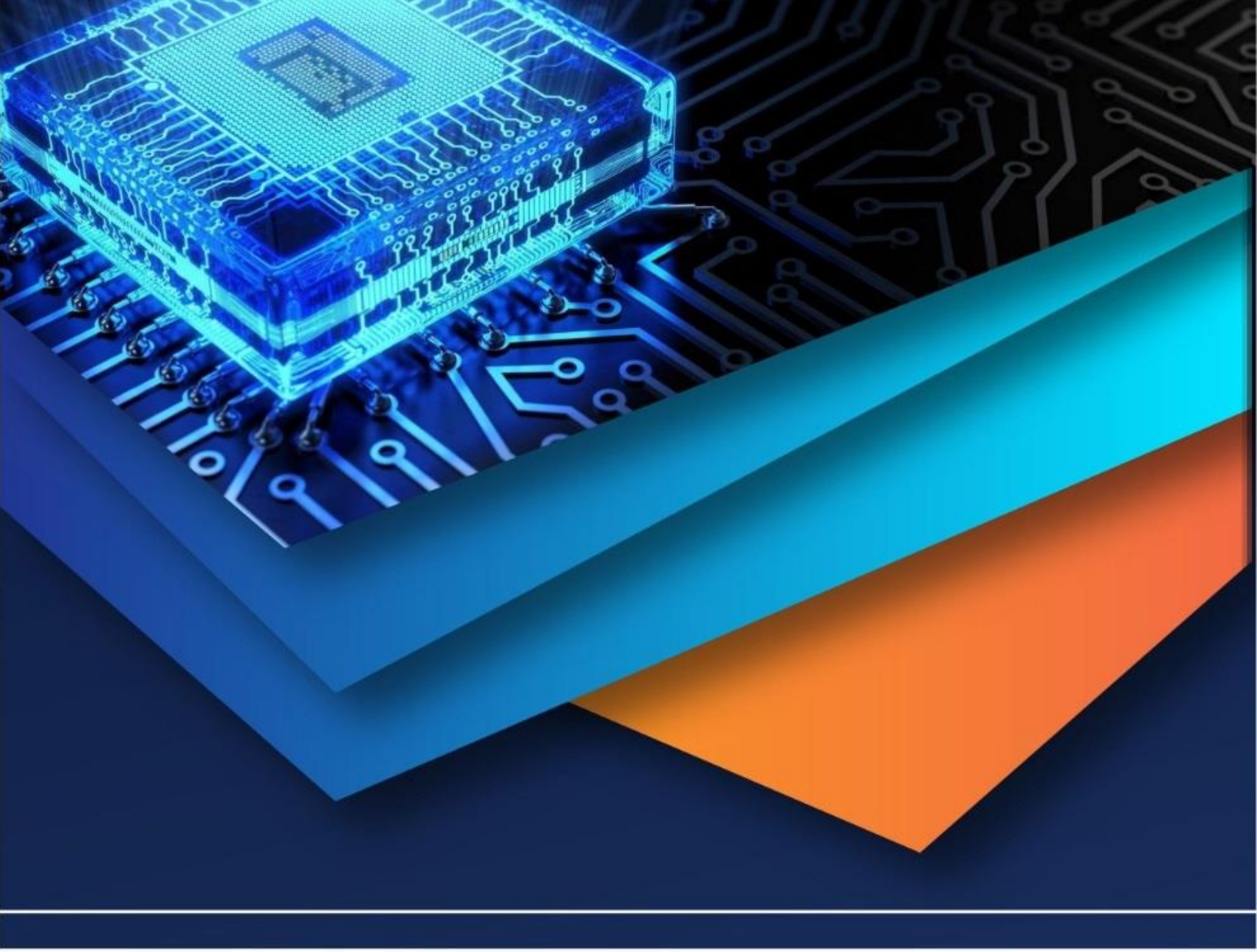
The results obtained through extensive experimentation validate the system's ability to accurately distinguish between legitimate and malicious accounts. Models such as Support Vector Machines have demonstrated high precision, recall, and AUC values, confirming the robustness of the selected features and the effectiveness of the training process. Visual evaluation using confusion matrices and ROC curves further reinforces the framework's suitability for large-scale, automated deployment. The system achieves a well- balanced trade-off between accuracy and interpretability,makingitvaluablenotonly for researchers but also for cybersecurity professionals and social media platforms seeking reliable detection mechanisms.

Looking ahead, the framework can be further strengthened through several enhancements. Integrating deep learning modelssuchasRecurrentNeuralNetworks (RNNs) could improve performance on more complex text and temporal patterns. Real-time detection capability can be introduced using streaming data pipelines, enabling instant flagging of suspicious activity. Incorporating Natural Language Processing (NLP) for semantic analysis of user-generatedcontentwouldaddanotherlayer of insight, especially in identifying harmfulormisleadinginformation.Finally, cross-platform integration and support for multilingual analysis would expand the system's applicability in diverse global contexts, making it a holistic solution for combating digital misinformation and abuse.

REFERENCES

[1] Benevenuto, F., Magno, G., Rodrigues, T., &Almeida,V.(2010).DetectingspammersonTwitter. Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (CEAS).
[2] Yang, C., Harkreader, R., & Gu, G. (2011). Die free or live hard? Empirical evaluation and new design for fighting evolving Twitter spammers. Recent Advances in Intrusion Detection (RAID).
[3] Lee,K.,Eoff,B.D.,&Caverlee,J.(2011).Seven monthswiththedevils:Along-termstudyofcontent polluters on Twitter. ICWSM.
[4] Stringhini,G.,Kruegel,C.,&Vigna,G.(2010). Detecting spammers on social networks. ACSAC.

[5] Almaatouq, A., Radaelli, L., Pentland, A., & Shmueli, E. (2016). Are you your friends' friend? Poor perception of friendship ties limits the ability to promote behavioral change. PLoS ONE.
[6] Echeverría, J., & Zhou, S. (2017). Discovery of the Twitter botnet "Star Wars". arXiv preprint arXiv:1701.02405.
[7] Chu,Z.,Gianvecchio,S.,Wang,H.,&Jajodia,S.(2012).DetectingautomationofTwitteraccounts: Areyouahuman,bot,orcyborg?IEEETransactions on Dependable and Secure Computing, 9(6), 811- 824.
[8] Dickerson,J.P.,Kagan,V.,&Subrahmanian,V.S.(2014).UsingsentimenttodetectbotsonTwitter: Are humans more opinionated than bots? IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).
[9] Ahmed, F., &Abulaish, M. (2013). A generic statisticalapproachforspamdetectioninOnline Social Networks. Computer Communications, 36(10), 1120–1129.
[10] Ferrara,E.,Varol,O.,Davis,C.,Menczer,F.,& Flammini, A. (2016). The rise of social bots. Communications of the ACM, 59(7), 96–104. Social Networks. Computer Communications, 36(10), 1120–1129.
[11] Ferrara,E.,Varol,O.,Davis,C.,Menczer,F.,& Flammini, A. (2016). The rise of social bots. Communications of the ACM, 59(7), 96–104.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY