National Conference on Information Technology & Digital Application (NCITSA-2025)
International Journal of Advanced Engineering Research and Science (IJAERS) Special Issue
https://dx.doi.org/10.22161/ijaers.ncitsa.2025.10 ISSN: 2349-6495(P) | 2456-1908(O)

Fake News Detection using Artificial Intelligence

Mr. Md. Rizwan Khan[1] , Yash Sharma[2], Somya Agrawal[3]

[1]Assistant Professor
Department of Information Technology,Jaipur Engineering College & Research Centre, Jaipur
[2]Student
Department of Information Technology, Jaipur Engineering College and Research Centre, Jaipur
Email: yashsharma.it25@gmail.com
[3]Student
Department of Information Technology, Jaipur Engineering College and Research Centre, India
Email: somyaagrawal.it25@gmail.com

Abstract—The Internet of Things (IoT) has become a significant part of everyday life, transforming how people interact with technology and the world around them. The number of connected devices has increased exponentially, leadingtounprecedentedconvenienceandefficiencyacrossvariousdomains,includinghealthcare,smartcities, agriculture, and industrial automation.Cyber threats targeting IoT systems have grown alongside the technology, highlighting the need for heightened awareness and robust security measures.analyze its potential impact across sectors, and emphasize the importance of addressing security concerns. Through a comprehensive analysis, the study underscores the dual nature of IoT—as a driver of innovation and a potential vector for cyber threats. The findings emphasize the necessity of informed policies, advanced technologies, and collaborative efforts to ensure that IoT benefits society while mitigating associated risks

I. INTRODUCTION

The Internet has become a cornerstone of modern life, enabling a wide range of activities in entertainment, business, education, and more. In India, the rapid growth of the internet has significantly reshaped industries, transforming the way businesses operate and how individuals engage with digital services. This digital revolution has not only facilitated access to global markets but also enhanced the quality of life by providing new opportunities for education, healthcare, and communication. From e-commerce platforms reaching millions of customers to online learning systems, the Internet has opened up avenues that were once unimaginable. However, this increasing reliance on the internet has introduced a new set of challenges,particularlytheriseofcybercrimes.With the growing number of devices connected to the internet and the expansion of digital services, cybercriminals have found new ways to exploit vulnerabilities. In India, the number of cyberattacks has escalated in recent years, with the country ranking as the third-largest source of malicious internet activity globally. Cybercrimes include a varietyofactivitiessuchasphishing,viruses, ransomware, data breaches, and denial-of-serviceattacks, all of which can cause significant harm to individuals, businesses, and even governmentinstitutions.

II. LITERATUREREVIEW

The challenge of detecting anomalous behavior in IoT networks has attracted significant attention in recent years. Early research largely employed statistical and linguistic methods, focusing on features such as signal frequency, protocol tags, and network traffic patterns. Classifiers such as Naive Bayes and Support Vector Machines (SVM) showed moderate success, offering interpretable results but limited accuracy in complex scenarios.

Subsequent research adopted more advanced techniques. Recurrent Neural Networks (RNNs), particularly LSTM networks, emerged as effective tools for modeling sequential data such as IoT device logs. CNNs were also explored for their ability to capture hierarchical features in signal or traffic data. More recently, transformer-based models like BERT have redefined sequence modeling tasks, including IoT anomaly detection. These models offer contextual embeddings that significantly improve classification performance.

Datasets such as TON_IoT, BoT-IoT, and UNSW-NB15 have become standard benchmarks for evaluating model accuracy.

https://dx.doi.org/10.22161/ijaers.ncitsa.2025.10 ISSN: 2349-6495(P) | 2456-1908(O)

Studies have also explored the incorporation of metadata—such as device type, manufacturer credibility, and timestamp logs—to improve prediction. Some researchers advocate hybrid models that combine traffic and contextual data, while others have explored the use of ensemble learning for greater robustness.

III. METHODOLOGY

3.1 : Dataset

This study utilizes the TON_IoT Dataset, which includes thousands of real and malicious IoT traffic samples from reputable and non-reputable sources. The dataset spans multiple domains, including smart homes, healthcare, and industrial control systems. Each entry contains both packet metadata and full payload logs, providing sufficient textual and numeric data for effective model training and evaluation.

3.2 : Preprocessing

Effective preprocessing is essential for ensuring high model accuracy. In our study, preprocessing involves removing HTML-like headers, special characters, and punctuation; converting all text to lowercase; tokenizing signal streams; removing redundant data; and normalizing traffic values to their base forms. Duplicate entries and irrelevant metadata are eliminated to maintain dataset integrity. This helps in reducing noise and focusing the learning models on meaningful patterns.

3.3 : Feature Extraction

For feature extraction, we employ both traditional and modern techniques:

• **TF-IDF (Term Frequency-Inverse Document Frequency)**: Used for quantifying the importance of command and control patterns in traffic relative to the dataset.

• **Signal Embeddings (Word2Vec, GloVe)**: Capture semantic relationships between device commands or payloads, offering richer context for deep learning models. These representations convert traffic logs into numerical vectors that serve as inputs to machine learning and deep learning models.

3.4 : Models Used

To evaluate the effectiveness of different AI approaches, we implement and compare the following models:

• **Naive Bayes**: A probabilistic classifier based on Bayes' theorem.

• **Logistic Regression**: A linear model suited for binary classification.

• **Random Forest**: An ensemble of decision trees for better generalization.

• **LSTM (Long Short-Term Memory)**: A type of RNN designed to handle long-range dependencies in IoT signal sequences.

• **BERT (Bidirectional Encoder Representations from Transformers)**: A transformer-based model pre-trained on large corpora and fine-tuned for sequence classification.

IV. IMPLEMENTATION

Implementation is carried out using Python and its scientific libraries. Traditional ML models are developed using Scikit-learn, while deep learning models utilize TensorFlow and Keras. BERT is implemented using HuggingFace Transformers, which provides pre-trained language models with high performance.

The dataset is divided into 80% training and 20% testing sets. Each model is trained using appropriate hyperparameters and evaluated on the test set. Cross-validation is used to ensure the robustness of results. Metrics such as accuracy, precision, recall, and F1-score are used for model comparison. Confusion matrices are employed to visualize model errors and better understand their strengths and weaknesses.

To handle the high dimensionality and complexity of IoT data, dimensionality reduction techniques such as PCA (Principal Component Analysis) were also explored during preprocessing to enhance model efficiency. Additionally, hyperparameter tuning was performed using grid search and random search methods to optimize each model's performance. The training process was conducted on GPU-enabled systems to reduce computation time, especially for deep learning models like LSTM.

III. RESULTSANDDISCUSSION

The models were compared on their performance, and the results are summarized below:

Model	Accuracy	Precision	Recall	F1-Score
Naive Bayes	88.2%	87.5%	89.0%	88.2%
Logistic	91.4%	91.0%	91.8%	91.4%

https://dx.doi.org/10.22161/ijaers.ncitsa.2025.10 ISSN: 2349-6495(P) | 2456-1908(O)

Model	Accuracy	Precision	Recall	F1-Score
Regression				
Random Forest	92.3%	92.0%	93.1%	92.5%
LSTM	94.5%	94.0%	95.2%	94.6%
BERT	96.8%	96.5%	97.1%	96.8%

The findings reveal that deep learning models outperform traditional machine learning algorithms in IoT anomaly detection. BERT, in particular, shows exceptional performance due to its ability to capture bidirectional context and semantic depth in sequential device or network data. While LSTM also performs well, BERT's architecture provides a more comprehensive understanding, though it demands greater computational resources.

Traditional models like Logistic Regression and Random Forest offer faster training times and require less computational power, making them suitable for IoT systems with resource constraints, such as edge devices. However, their performance, while respectable, falls short of deep learning approaches in terms of precision and recall.

These results suggest that model selection should be aligned with specific application requirements. For real-time, large-scale IoT environments, it is important to strike a balance between detection accuracy and system efficiency.

IV. CONCLUSION

This paper explored the application of AI-based methods for IoT anomaly detection. Our evaluation of various models—including Naive Bayes, Logistic Regression, Random Forest, LSTM, and BERT—demonstrates that modern deep learning models significantly outperform traditional techniques. BERT emerged as the most effective model, achieving the highest scores across all evaluation metrics.

The implications of this study are significant for IoT infrastructure providers, system administrators, and AI researchers. Integrating these models into real-time monitoring systems could enhance the detection of anomalies, faults, or potential cyber threats across IoT networks. However, the trade-off between model complexity and performance must be carefully considered, especially in resource-constrained environments like edge devices.

https://dx.doi.org/10.22161/ijaers.ncitsa.2025.10 ISSN: 2349-6495(P) | 2456-1908(O)

Futureresearchcanexploremulti-modaldetectionmethods that combine text, images, and metadata. Real-time analysis, multilingual datasets, and ethical considerations such as model transparency and fairness should also be prioritized.

ACKNOWLEDGEMENTS

We extend our gratitude to the open-source communities and IoT dataset providers whose resources made this research possible. Their contributions are foundational to innovation in AI-driven IoT solutions.

REFERENCES

1. Atzori,L.,Iera,A.,&Morabito,G.(2010): The Internet of Things: A Survey. *Computer Networks, 54(15), 2787-2805.*

(AcomprehensivesurveyonIoTapplications and research directions.)

2. Evans, D. (2011): The Internet of Things: How the Next Evolution of the Internet Is ChangingEverything.*CiscoInternetBusiness Solutions Group (IBSG).*

(DiscussesthepotentialofIoTintransforming industries and societies.)

3. Perera, C., Zaslavsky, A., Christen, P., & Georgakopoulos, D. (2014): Context-Aware ComputingfortheInternetofThings:ASurvey. *IEEE Communications Surveys & Tutorials, 16(1), 414-454.*

(Focusesoncontext-awarecomputingasa critical aspect of IoT.)

4. Bandyopadhyay, D., & Sen, J. (2011): InternetofThings:ApplicationsandChallenges in Technology and Standardization. *Wireless Personal Communications, 58, 49-69. (Provides a detailed discussion on IoT applications and standardization challenges.)*
5. Zanella, A., Bui, N., Castellani, A., Vangelista,L.,&Zorzi,M.(2014):Internetof Things for Smart Cities. *IEEE Internet of Things Journal, 1(1), 22-32.*

(ExploresIoTapplicationsforthedevelopment of smart cities.)

6. Miorandi, D., Sicari, S., De Pellegrini, F., & Chlamtac, I. (2012): Internet of Things: Vision,Applications,andResearchChallenges. *Ad Hoc Networks, 10(7), 1497-1516. (Highlightsthechallengesandfuturedirections in IoT research.)*