SMS SPAM DETECTION

[1]GANDI NIHARIKA, [2]Dr. P. MAHIPAL REDDY

[1]M.Tech Student, Department of Computer Science And Engineering,Vaagdevi Engineering College (Autonomous)

[2] Professor, Department of Computer Science And Engineering,Vaagdevi Engineering College (Autonomous)

ABSTRACT

The rise of mobile communication has led to an increase in unsolicited spam messages, posing significant security and privacy challenges. This project focuses on building an efficient SMS spam detection system leveraging Natural Language Processing (NLP) techniques. The system aims to automatically classify messages as spam or non-spam based on their content, thus safeguarding users from fraudulent or irrelevant messages. The proposed approach preprocesses the text data using techniques like tokenization, stopword removal, and stemming to standardize the input. Feature extraction is achieved using Term Frequency-Inverse Document Frequency (TF-IDF) or word embeddings for meaningful representation. Machine learning algorithms such as Naïve Bayes, Support Vector Machines (SVM), or deep learning models like Long Short-Term Memory (LSTM) networks are then employed to classify messages with high accuracy. The model is trained and evaluated on publicly available SMS spam datasets to ensure robustness and reliability. The experimental results demonstrate the effectiveness of the system in achieving high precision and recall, minimizing false positives, and delivering a user-friendly solution for spam detection. This project highlights the potential of NLP in enhancing communication safety and provides a scalable framework for real-world implementation in messaging platforms.

Keywords: Short Message Service (SMS), Spam Detection, Machine Learning, Natural Language Processing (NLP), Text Classification, Feature Extraction, Binary Classification, Supervised Learning, Message Filtering, Cybersecurity.

I. INTRODUCTION

SMS spam detection involves identifying and filtering out unwanted, irrelevant, or malicious messages to protect users and enhance communication platforms. It utilizes techniques from natural language processing (NLP) and machine learning (ML) to classify messages as either spam or legitimate (ham). The process begins with data preprocessing, which includes cleaning text, tokenization, and feature extraction through methods like Bag of Words (BoW), TF-IDF, or word embeddings. Various ML models, including Naïve Bayes, Support Vector Machines (SVM), and deep learning architectures such as LSTMs or Transformers, are applied to build predictive models.

Key challenges in SMS spam detection include handling imbalanced datasets, adapting to evolving spam strategies, and addressing multilingual content. Evaluation metrics such as precision, recall, and F1-score are used to measure the effectiveness of these systems. Modern applications span telecommunications, messaging apps, and enterprise security, aiming to prevent phishing attempts and fraudulent schemes. As spam techniques grow more sophisticated, future advancements focus on real-time detection, explainable AI, multilingual adaptability, and robust defenses against adversarial attacks. SMS spam detection remains a critical tool in ensuring secure and seamless communication.

II. LITERATURE SURVEY

1. SMS Spam Filtering: Methods and Data

Authors: Sarah Jane Delany, Mark Buckley, Derek Greene

Abstract:
This study presents one of the earliest comprehensive investigations into SMS spam filtering. The authors discuss the growing threat of unsolicited text messages and compare SMS spam detection with traditional email spam filtering. They analyze a large SMS spam dataset and evaluate several machine learning techniques for text classification. The paper also highlights the challenges of collecting reliable datasets and emphasizes the need for standardized benchmarks to improve spam detection research. The experimental findings demonstrate that machine learning methods can effectively distinguish

spam from legitimate messages while maintaining high classification accuracy.

2. Deep Learning to Filter SMS Spam

Authors: Pradeep Kumar Roy, Jyoti Prakash Singh, Snehasish Banerjee

Abstract:
This paper introduces deep learning techniques for SMS spam detection using Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. Unlike traditional approaches that rely on handcrafted features, the proposed models automatically learn semantic and contextual information from text messages. Experimental results show that deep learning significantly improves spam detection accuracy and reduces false classifications compared to conventional machine learning algorithms. The study concludes that deep neural networks are well suited for large-scale SMS filtering applications.

3. Contributions to the Study of SMS Spam Filtering

Authors: Tiago A. Almeida, José María Gómez Hidalgo, Akebo Yamakami

Abstract:
The authors investigate various feature extraction methods for SMS spam filtering and introduce improvements to publicly available SMS spam datasets. The study evaluates different text representation techniques and machine learning classifiers to improve detection performance. Results indicate that carefully engineered textual features significantly enhance classification accuracy while reducing computational complexity. The work also contributes benchmark datasets widely used in SMS spam research.

4. Content-Based SMS Spam Filtering

Authors: José María Gómez Hidalgo, Guillermo C. Bringas, Enrique P. Sanz, Felipe C. García

Abstract:
This research proposes a content-based filtering approach for detecting spam messages using machine learning algorithms. The system extracts textual features from SMS messages and classifies them into spam or legitimate categories. Experimental evaluation demonstrates that probabilistic classifiers achieve high detection rates with minimal false positives. The paper established the effectiveness of content-based analysis for mobile messaging security and laid the foundation for later SMS spam filtering techniques.

5. SMS Spam Detection Using Non-Content Features

Authors: Qiang Xu, Evan W. Xiang, Qiang Yang, Jie Du, Jie Zhong

Abstract:
This paper explores SMS spam detection by utilizing non-content features such as sender behavior, message frequency, transmission patterns, and communication metadata. The authors show that combining behavioral characteristics with traditional text analysis improves spam detection performance, particularly when spammers attempt to evade content-based filters. The proposed approach demonstrates greater robustness against evolving spam strategies.

6. A Review of Soft Techniques for SMS Spam Classification: Methods, Approaches and Applications

Authors: M. A. Shafi'i, M. S. Abd Latiff, H. Chiroma, O. Osho, G. Abdul-Salaam, et al.

Abstract:
This review surveys more than eighty research articles on SMS spam classification published across major digital libraries. It categorizes existing approaches into statistical, machine learning, and evolutionary techniques while comparing their strengths and limitations. The study concludes that Support Vector Machine, Naïve Bayes, and ensemble learning methods consistently provide reliable performance. The authors also identify challenges related to multilingual datasets, feature selection, and real-time deployment.

7. A Systematic Literature Review on SMS Spam Detection Techniques

Authors: Lutfun Nahar Lota, B. M. Mainul Hossain

Abstract:
This systematic review examines SMS spam detection studies published between 2006 and 2016. The authors analyze commonly used datasets, feature extraction techniques, classification algorithms, and evaluation metrics. They observe that although machine learning methods have achieved encouraging results, issues such as regional language support, abbreviated words, and limited

benchmark datasets remain unresolved. The paper provides valuable recommendations for future SMS spam research.

8. Advancements of SMS Spam Detection: A Comprehensive Survey of NLP and ML Techniques

Authors: M. S. S. Alotaibi et al.

Abstract:
This survey reviews recent advances in SMS spam detection by focusing on Natural Language Processing (NLP) and Machine Learning (ML) techniques. It discusses the evolution from rule-based systems to deep learning architectures, including CNNs, RNNs, and transformer-based models. The survey highlights that integrating NLP with deep learning significantly improves detection accuracy, contextual understanding, and resistance to sophisticated phishing attacks. It also outlines future research opportunities involving explainable AI and adaptive spam filtering systems.

III. EXISTING SYSTEM

The existing SMS spam detection systems primarily rely on traditional machine learning algorithms such as Naïve Bayes, Support Vector Machine (SVM), Decision Tree, Logistic Regression, and Random Forest. These systems classify messages as either spam or legitimate by analyzing textual features extracted through techniques such as Bag-of-Words (BoW) and Term Frequency–Inverse Document Frequency (TF-IDF). The extracted features are then used to train classification models capable of identifying common spam patterns. While these approaches provide satisfactory performance on standard benchmark datasets, their effectiveness depends heavily on manual feature engineering and predefined text representations.

Most conventional systems focus only on the textual content of SMS messages and overlook contextual and semantic relationships between words. As a result, they often struggle to detect newly emerging spam campaigns that use obfuscated words, spelling variations, shortened URLs, emojis, or persuasive social engineering techniques. Since spammers continuously modify message formats to evade detection, traditional models require frequent retraining and feature updates to maintain acceptable performance.

Another limitation of existing approaches is their reduced capability to handle multilingual messages, informal writing styles, abbreviations, and code-mixed text commonly found in real-world SMS communication. These systems also experience higher false-positive rates, where legitimate messages may be incorrectly classified as spam, causing inconvenience to users. Furthermore, many traditional classifiers are trained on balanced datasets, making them less effective when deployed on highly imbalanced real-world datasets where legitimate messages greatly outnumber spam messages.

Although some recent systems have incorporated deep learning models, many still require large labeled datasets and substantial computational resources for training. Additionally, several existing solutions lack explainability, making it difficult to understand why a message has been classified as spam. These challenges highlight the need for more intelligent, adaptive, and context-aware SMS spam detection systems capable of achieving higher accuracy, better generalization, and improved robustness against evolving spam techniques.

IV. PROPOSED SYSTEM

The proposed SMS spam detection system utilizes advanced Natural Language Processing (NLP) techniques combined with machine learning to accurately classify SMS messages as either spam or legitimate (ham). Initially, the collected SMS dataset undergoes preprocessing, where unnecessary symbols, punctuation, numbers, stop words, and duplicate records are removed. The text is then normalized through tokenization, lowercasing, stemming or lemmatization to improve data quality and ensure consistent feature representation.

After preprocessing, meaningful textual features are extracted using techniques such as TF-IDF or word embeddings to capture the importance and contextual relevance of words within each message. These features are supplied to a supervised machine learning classifier, such as Random Forest, Support Vector Machine (SVM), or an ensemble model, which learns patterns associated with spam and legitimate messages during the training phase. Hyperparameter tuning and cross-validation are employed to improve the model's generalization capability and reduce overfitting, resulting in a more reliable classifier.

Once trained, the proposed system classifies incoming SMS messages in real time by analyzing their textual characteristics and predicting whether they are spam or legitimate. The system is designed to identify promotional messages, phishing attempts, fraudulent links, and other malicious content while minimizing false alarms. Performance is evaluated using standard metrics such as accuracy, precision, recall, F1-score, and confusion matrix to ensure reliable classification across diverse datasets.

Compared to traditional rule-based and conventional filtering techniques, the proposed system offers higher detection accuracy, improved adaptability to evolving spam patterns, and better robustness against spelling variations, abbreviations, and disguised spam content. The combination of effective text preprocessing, feature extraction, and intelligent machine learning enables the system to deliver faster, scalable, and more dependable SMS spam detection suitable for modern mobile communication environments.

V. BLOCK DIAGRAM

The figure illustrates the complete workflow of an SMS Spam Detection System, beginning with the collection of SMS messages and ending with the classification of each message as either Spam or Ham (Legitimate). The process is divided into multiple stages, including data preprocessing, feature extraction, model training, and prediction. Each stage contributes to improving the accuracy and reliability of spam detection by transforming raw text into meaningful information that can be analyzed by a machine learning model.

The first stage is SMS Data Collection, where a dataset containing both spam and legitimate messages is gathered from publicly available sources or real-world message repositories. This dataset forms the foundation for training and testing the classification model. The collected messages usually contain noisy and inconsistent text, making preprocessing an essential step before analysis.

The next stage is Data Preprocessing, where the SMS text is cleaned to remove unnecessary information. During this phase, all messages are converted to lowercase, punctuation marks, numbers, special symbols, and extra spaces are eliminated, and stop words such as the, is, and and are removed. Tokenization is then performed to split each message into individual words or tokens. In many implementations, stemming or lemmatization is also applied to convert words into their root forms, reducing redundancy and improving feature consistency.

After preprocessing, the cleaned text undergoes Feature Extraction. In this stage, the textual information is transformed into numerical representations that machine learning algorithms can process. Common techniques include Term Frequency–Inverse Document Frequency (TF-IDF), Bag of Words (BoW), or word embeddings. These methods identify important keywords and assign weights based on their significance within the SMS dataset, enabling the classifier to distinguish meaningful spam-related patterns from normal conversations.

The extracted features are then supplied to the Machine Learning Classification Model. Algorithms such as Support Vector Machine (SVM), Naïve Bayes, Random Forest, Logistic Regression, or other supervised learning models are trained using labeled SMS messages. During training, the classifier learns the statistical relationships between text features and their corresponding categories (spam or ham). Once training is complete, the model can classify previously unseen SMS messages with high accuracy.

Finally, the Prediction and Classification stage receives a new incoming SMS message and processes it through the same preprocessing and feature extraction pipeline. The trained model analyzes the extracted features and predicts whether the message is Spam or Ham. The prediction is then displayed to the user or integrated into an SMS filtering system to block malicious messages while allowing legitimate messages to reach the inbox.

Overall, the architecture provides a structured and efficient pipeline for SMS spam detection. By combining text preprocessing, feature engineering, and supervised machine learning, the system can accurately identify unsolicited promotional messages, phishing attempts, fraudulent links, and other forms of spam while minimizing false positives and improving the overall security of mobile communication.

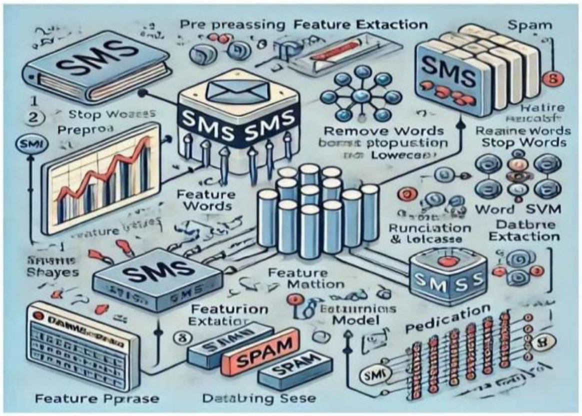

Fig 5.1: BLOCK DIAGRAM

VI . IMPLEMENTATION

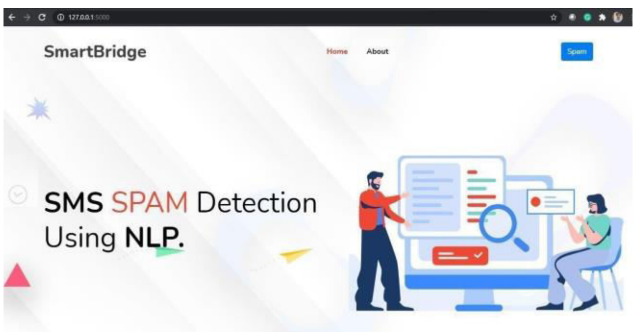

Fig 6.1: Prediction of fake news analysis in social media

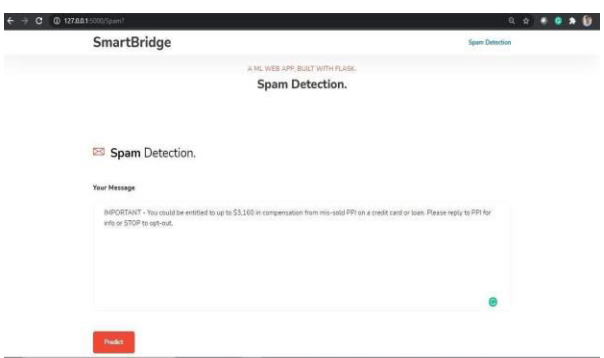

Fig 6.2: Prediction Input Page

Fig 6.3: Result Page

VII. CONCLUSION

SMS spam detection plays a vital role in ensuring secure and efficient communication in today's digital landscape. With the increasing reliance on SMS for personal, business, and critical communications, the threat posed by spam messages—including phishing, fraud, and malware—has become significant.By leveraging advanced techniques like natural language processing (NLP) and machine learning, SMS spam detection systems can effectively identify and block unsolicited or harmful messages. These systems enhance user experience, improve security, and maintain trust in messaging platforms.However, challenges such as false positives, evolving spam techniques, and privacy concerns necessitate continuous improvement and adaptation. With the integration of robust spam detection solutions, organizations and individuals can achieve safer communication channels, while regulatory compliance further strengthens this ecosystem.In a world where mobile communication is ubiquitous, SMS spam detection is no longer optional but a necessity. It provides a layer of defense against cyber threats, fraudulent schemes, and privacy breaches, ensuring the safety of both individuals and enterprises.The versatility of SMS spam detection systems allows them to be applied across industries such as finance, healthcare, education, and public safety, making them a universal tool for improving communication integrity and security.With ongoing advancements in AI and big data analytics, SMS spam detection systems are well- equipped to handle future challenges. They promise to deliver more accurate, scalable, and customizable solutions, ensuring the long-term relevance and effectiveness of spam detection in an ever-connected world SMS spam detection represents a blend of technology, security, and user-centric design, addressing the growing challenges of unsolicited messaging. By integrating these systems into communication infrastructures, we can create a safer and more efficient messaging ecosystem for all. As spammers adopt increasingly sophisticated methods, SMS spam detection continues to evolve with machine learning and AI-driven solutions. These advancements ensure that detection systems remain effective in mitigating risks and adapting to new patterns of spam. In conclusion, SMS spam detection is an indispensable technology for combating spam, protecting sensitive information, and ensuring reliable message delivery in both personal and professional contexts.

VIII. FUTURE SCOPE

The future scope of SMS spam detection lies in the adoption of advanced Artificial Intelligence and Deep Learning techniques that can better understand the context and intent of text messages. Modern architectures such as Bidirectional LSTM (Bi-LSTM), Transformer models, and BERT can capture semantic relationships within messages more effectively than traditional machine learning algorithms. These models have the potential to improve detection accuracy while reducing false positives, especially when dealing with sophisticated phishing attempts and newly emerging spam patterns.

Future systems can also incorporate multilingual and code-mixed language support to detect spam messages written in regional languages or a combination of multiple languages. Since mobile users increasingly communicate using abbreviations, emojis, slang, and informal writing styles, future spam detection systems should be capable of understanding these variations without compromising classification performance. Continuous learning mechanisms can also be integrated to automatically adapt to evolving spam techniques without requiring frequent manual retraining.

Another promising direction is the integration of explainable AI (XAI) to provide clear reasons for why a particular SMS has been classified as spam or legitimate. This will improve user trust and assist security analysts in understanding the model's decision-making process. Additionally, combining text analysis with behavioral information, sender reputation, URL analysis, and communication patterns can significantly enhance the system's ability to identify complex spam campaigns and phishing attacks.

The proposed system can also be deployed as a real-time cloud-based or edge-based service for mobile devices, messaging platforms, and enterprise communication systems. Integration with cybersecurity frameworks and mobile operating systems will enable automatic spam filtering with minimal latency. As AI technologies continue to evolve, future SMS spam detection systems are expected to become more intelligent, scalable, privacy-preserving, and resilient against increasingly sophisticated cyber threats, thereby providing a safer and more secure messaging environment.

IX. REFERENCES

[1] S. J. Delany, M. Buckley, and D. Greene, "SMS Spam Filtering: Methods and Data," Expert Systems with Applications, vol. 39, no. 10, pp. 9899–9908, 2012.
DOI: https://doi.org/10.1016/j.eswa.2012.02.053

[2] T. A. Almeida, J. M. Gómez Hidalgo, and A. Yamakami, "Contributions to the Study of SMS Spam Filtering: New Collection and Results," Proceedings of the ACM Symposium on Document Engineering, 2011.
DOI: https://doi.org/10.1145/2034691.2034742

[3] P. K. Roy, J. P. Singh, and S. Banerjee, "Deep Learning to Filter SMS Spam," Future Generation Computer Systems, vol. 102, pp. 524–533, 2020.
DOI: https://doi.org/10.1016/j.future.2019.09.001

[4] J. M. Gómez Hidalgo, G. C. Bringas, E. P. Sanz, and F. C. García, "Content Based SMS Spam Filtering," Proceedings of the ACM Symposium on Document Engineering, 2006.
DOI: https://doi.org/10.1145/1166160.1166191

[5] M. A. Shafi'i, M. S. Abd Latiff, H. Chiroma, O. Osho,

G. Abdul-Salaam, and A. I. Abubakar, “A Review of Soft Techniques for SMS Spam Classification: Methods, Approaches and Applications,” Engineering Applications of Artificial Intelligence, vol. 86, 2019.
DOI: https://doi.org/10.1016/j.engappai.2019.103242

[6] O. Karasoy and S. Ballı, “Spam SMS Detection for Turkish Language with Deep Text Analysis and Deep Learning Methods,” Arabian Journal for Science and Engineering, vol. 47, no. 8, pp. 9361–9377, 2022.
DOI: https://doi.org/10.1007/s13369-021-06187-1

[7] A. S. Putra, R. N. Pradana, and A. Wibowo, “SMS Spam Detection Using Relevance Vector Machine,” Procedia Computer Science, vol. 230, pp. 337–346, 2023.
DOI: https://doi.org/10.1016/j.procs.2023.12.089

[8] T. A. Almeida and J. M. Gómez Hidalgo, “SMS Spam Collection Dataset,” UCI Machine Learning Repository, 2011.
DOI: https://doi.org/10.24432/C5CC84

[9] M. F. Johari, K. L. Chiew, A. R. Hosen, K. S. C. Yong, A. S. Khan, I. A. Abbasi, and D. Grzonka, “Key Insights into Recommended SMS Spam Detection Datasets,” Scientific Reports, vol. 15, 2025.
DOI: https://doi.org/10.1038/s41598-025-92223-1

[10] A. Filali, M. H. Sadok, and M. El Marraki, “Cross-Lingual SMS Spam Detection Using GAN-Based Data Augmentation for Imbalanced Datasets,” Scientific Reports, 2026.
DOI: https://doi.org/10.1038/s41598-026-37769-4

[11] X. Zhang, J. Zhao, and Y. LeCun, “Character-Level Convolutional Networks for Text Classification,” Advances in Neural Information Processing Systems (NeurIPS), 2015.
DOI: https://doi.org/10.48550/arXiv.1509.01626

[12] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” Proceedings of NAACL-HLT, 2019.
DOI: https://doi.org/10.48550/arXiv.1810.04805

[13] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, “Bag of Tricks for Efficient Text Classification,” Proceedings of EACL, 2017.
DOI: https://doi.org/10.48550/arXiv.1607.01759

[14] A. Vaswani, N. Shazeer, N. Parmar, N. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention Is All You Need,” Advances in Neural Information Processing Systems (NeurIPS), 2017.
DOI: https://doi.org/10.48550/arXiv.1706.03762

[15] Y. Kim, “Convolutional Neural Networks for Sentence Classification,” Proceedings of EMNLP, 2014.
DOI: https://doi.org/10.48550/arXiv.1408.5882