SMS SPAM DETECTION

[1] P. Yamini Chouhan, [2] Uyyala Keerthana, [3] V Rakesh, [4] V.Mena Kiran
[1]AssistantProfessor, [234]Students
Department of Computer Engineering(Software Engineering)
Siddhartha Institute of Technology & Sciences, Narapally
yaminichouhan_cse@siddhartha.co.in, 23tq1a5615@siddhartha.co.in, 23tq1a5627@siddhartha.co.in, 23tq1a5616@siddhartha.co.in

Abstract

SMS spam has emerged as a significant challenge, posing risks to user privacy, security, and overall communication efficiency through the rapid spread of unsolicited and often malicious messages. Traditional rule-based filtering techniques are no longer sufficient to handle the dynamic and evolving nature of spam patterns. This project focuses on developing an efficient and reliable machine learning-based system to automatically classify SMS messages as either "spam" or "ham" (legitimate).

The model is built using the widely recognized UCI SMS Spam Collection dataset and incorporates a complete Natural Language Processing (NLP) pipeline. This includes data preprocessing steps such as text cleaning, tokenization, stop-word removal, and stemming to enhance the quality of textual data. Feature extraction is performed using TF-IDF vectorization to convert text into meaningful numerical representations.

A Support Vector Machine (SVM) classifier with a sigmoid kernel is employed for training and prediction.

I. Introduction

In today's digital era, mobile communication has become an integral part of daily life, with Short Message Service (SMS) being one of the most widely used and accessible forms of communication. Its simplicity, speed, and cost-effectiveness make it a preferred medium for personal, commercial, and organizational interactions. However, the increasing popularity of SMS has also led to a significant rise in spam messages, which negatively impact user experience and compromise security.

SMS spam refers to unsolicited and often harmful messages sent in bulk to users without their consent. These messages may include advertisements, fraudulent schemes, phishing links (commonly known as smishing), or malware intended to steal sensitive information such as passwords, banking details, or personal data. As mobile usage continues to grow globally, the volume and sophistication of spam messages are also increasing, making it a serious concern for both individuals and organizations.

Traditional spam filtering techniques, primarily designed for email systems, rely on static rules and keyword-based approaches. However, these methods are not well-suited for SMS data due to its unique characteristics, such as shorter message length, informal language, abbreviations, and frequent use of slang. As a result, there is a need for more advanced and adaptive approaches that can effectively identify and filter spam messages.

II. Literacture Survey

Research in SMS spam detection has been extensively explored using various machine learning and deep learning techniques. Early studies primarily focused on traditional machine learning algorithms such as Naive Bayes and Support Vector Machines (SVM), which proved to be highly effective for text classification tasks due to their ability to handle high-dimensional data efficiently. These methods became the foundation for many spam detection systems because of their simplicity, speed, and relatively high accuracy.

A significant contribution to this field was made by Almeida et al. (2011), who introduced the widely used *UCI SMS Spam Collection dataset*. This dataset has become a benchmark for evaluating SMS spam detection models and has been utilized in over 80 research studies. Their work provided baseline performance metrics for several classification algorithms, enabling researchers to compare and improve upon existing models.

In recent years, advancements in deep learning have led to the development of more sophisticated models for spam detection. Techniques such as Long Short-Term Memory (LSTM) networks and transformer-based models like BERT (Bidirectional Encoder Representations from Transformers) have shown remarkable performance improvements. These models are capable of capturing contextual information, semantic relationships, and sequential dependencies within text, leading to higher accuracy rates—often exceeding 99%, as reported in studies like Shen et al. (2025).

Despite these advancements, several challenges remain. One of the most critical issues highlighted in the literature is data imbalance, where spam messages are significantly fewer than legitimate (ham) messages.

III. System Analysis

The SMS spam detection system is designed to identify and classify messages as spam or ham using machine learning techniques. The existing systems rely on rule-based filtering, which is ineffective against evolving spam patterns and informal text used in SMS. These systems often result in high false positives and lack adaptability. To overcome these limitations, the proposed system uses Natural Language Processing (NLP) and machine learning algorithms. The system preprocesses text through cleaning, tokenization, and stop-word removal to improve data quality. Feature extraction is performed using TF-IDF to convert text into numerical form. A Support Vector Machine (SVM) classifier is trained to accurately classify messages. The system is capable of learning from data and adapting to new spam patterns. It provides high accuracy and efficiency in real-time message classification. Overall, the system is scalable, reliable, and suitable for practical applications in mobile communication.

Existing System

Traditional SMS spam detection systems are primarily based on rule-based filtering techniques and simple keyword matching methods. These systems identify spam messages using predefined rules such as detecting specific words (e.g., “win”, “free”,

"offer") or blacklisted numbers. They often rely on manually created filters and static databases, which require frequent updates to remain effective.

In many cases, these systems also use basic pattern recognition techniques, such as checking message frequency or identifying repeated content from unknown senders. However, they do not analyze the context or meaning of the message, making them less intelligent and adaptive. As spam techniques evolve, spammers use obfuscated words, symbols, and abbreviations to bypass these filters.

Disadvantages of Existing system

- Ineffective against evolving spam patterns and obfuscated text
- High false positive and false negative rates
- Lack of adaptability and learning capability
- Poor performance with informal language and abbreviations
- Cannot handle large-scale and real-time data efficiently

Proposed System

The proposed system utilizes Machine Learning (ML) and Natural Language Processing (NLP) techniques to build an intelligent SMS spam detection model. It classifies messages into spam or ham based on learned patterns from data. Unlike traditional systems, this approach is adaptive and improves its performance as more data is provided.

The system begins with data preprocessing, where SMS text is cleaned, tokenized, and normalized by removing stop words and applying stemming. The processed text is then converted into numerical form using TF-IDF (Term Frequency–Inverse Document Frequency), which highlights important words in the messages.

A Support Vector Machine (SVM) classifier with a sigmoid kernel is used to train the model, enabling it to effectively separate spam and ham messages. The system is capable of capturing patterns, context, and relationships within the text, resulting in high accuracy and precision.

Advantages of Proposed System

Hardware Requirements:

- Processor: Intel i3 or above
- RAM: Minimum 4 GB
- Storage: 10 GB free space

Software Requirements:

- Operating System: Windows/Linux/Mac
- Programming Language: Python
- Libraries: NumPy, Pandas, Scikit-learn, NLTK/Spacy
- Development Tools: VS Code / Jupyter Notebook

IV. Methodology

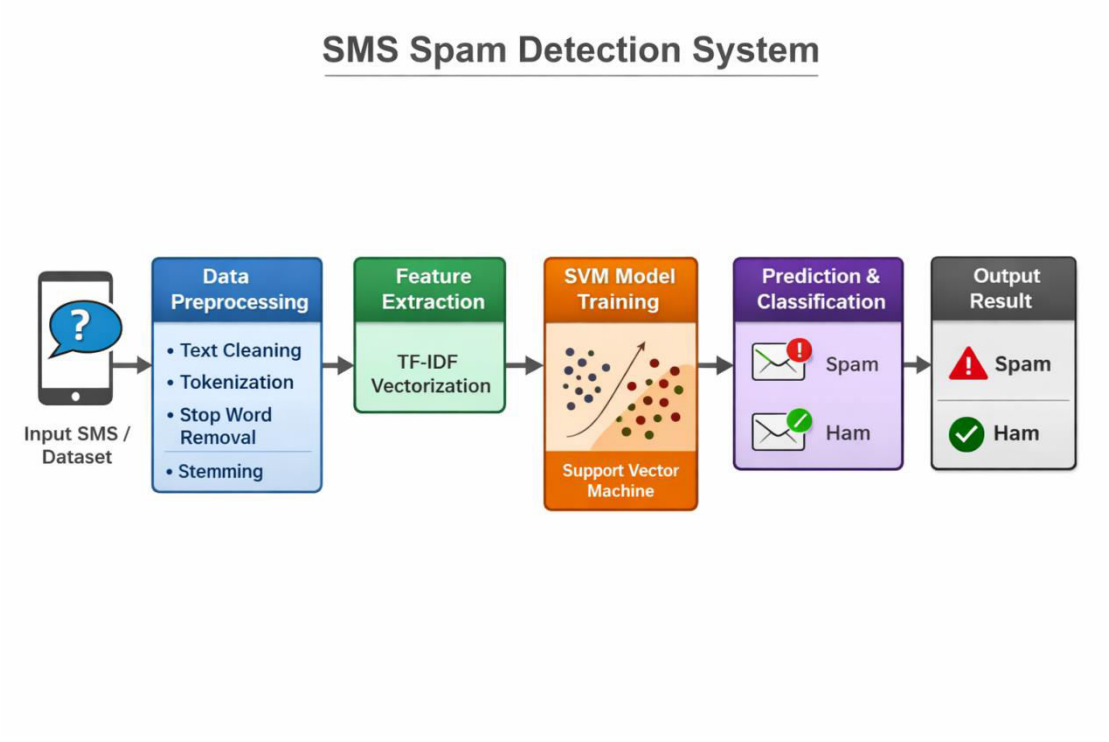
The methodology for SMS spam detection involves a sequence of steps to build an efficient and accurate classification model. Initially, the UCI SMS Spam Collection dataset is collected, which contains labeled messages as spam or ham. The next step is data preprocessing, where unwanted characters, punctuation, and numbers are removed. Text is converted to lowercase, followed by tokenization, stop-word removal, and stemming to normalize the data.

After preprocessing, feature extraction is performed using TF-IDF vectorization, which converts textual data into numerical form by assigning importance to words based on their frequency. The dataset is then divided into training and testing sets to evaluate the model's performance.

A Support Vector Machine (SVM) classifier with a sigmoid kernel is trained on the processed data. The model learns patterns and relationships between words to distinguish spam from legitimate messages. After training, the model is evaluated using metrics such as accuracy, precision, recall, and F1-score.

System Architecture

The system architecture defines the overall structure and workflow of the SMS spam detection system. It consists of multiple components that work together to process input messages and produce classification results.



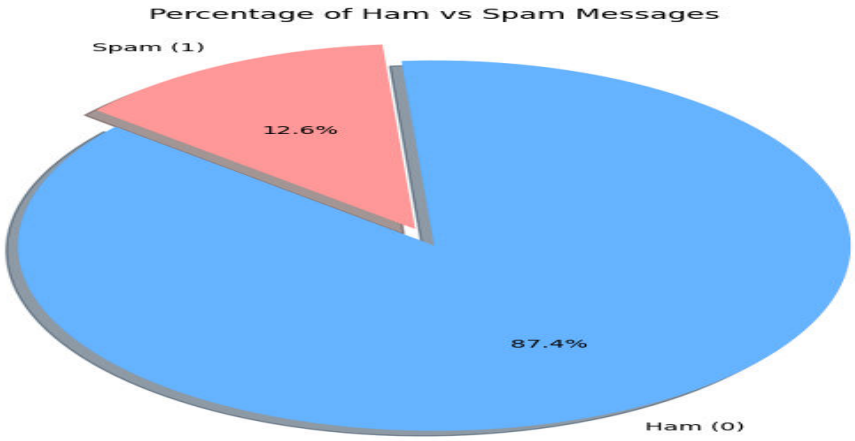
Architecture Description:

1. Input Layer:
   User provides an SMS message or dataset as input.
2. Data Preprocessing Module:
   Cleans and processes the text (removal of noise, tokenization, stop-word removal, stemming).

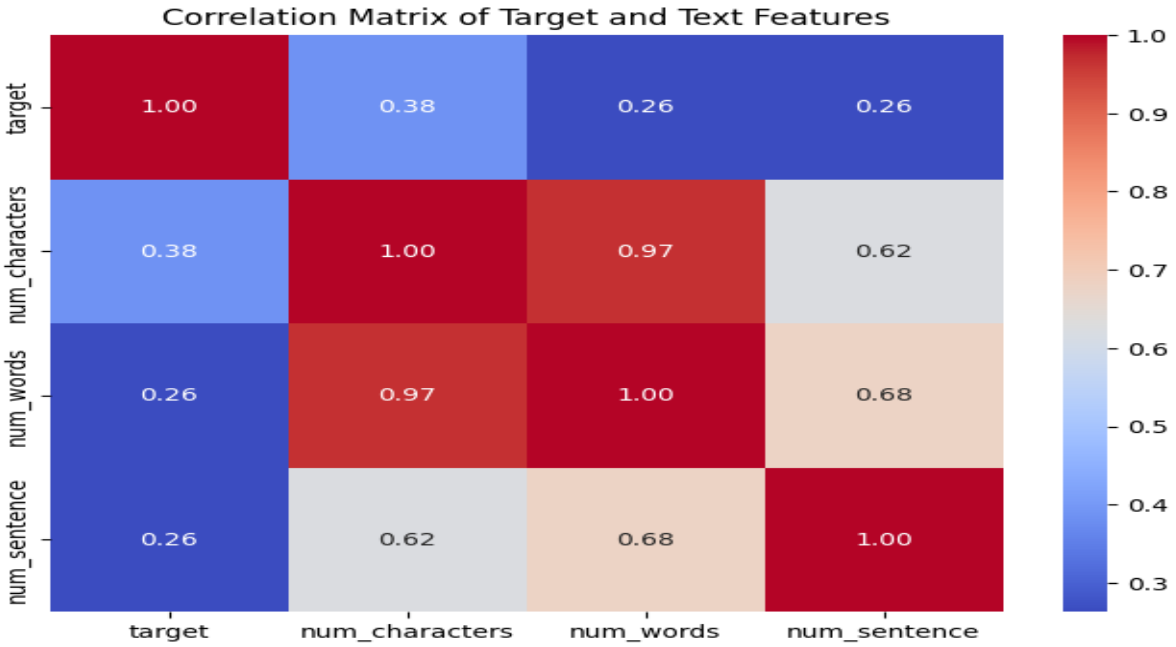
3. Feature Extraction Module:
   Converts processed text into numerical vectors using TF-IDF.
4. Model Training Module:
   Trains the SVM classifier using labeled dataset.
5. Prediction Module:
   Classifies messages as *spam* or *ham* based on trained model.
6. Output Layer:
   Displays the classification result to the user.

V. Result and Output

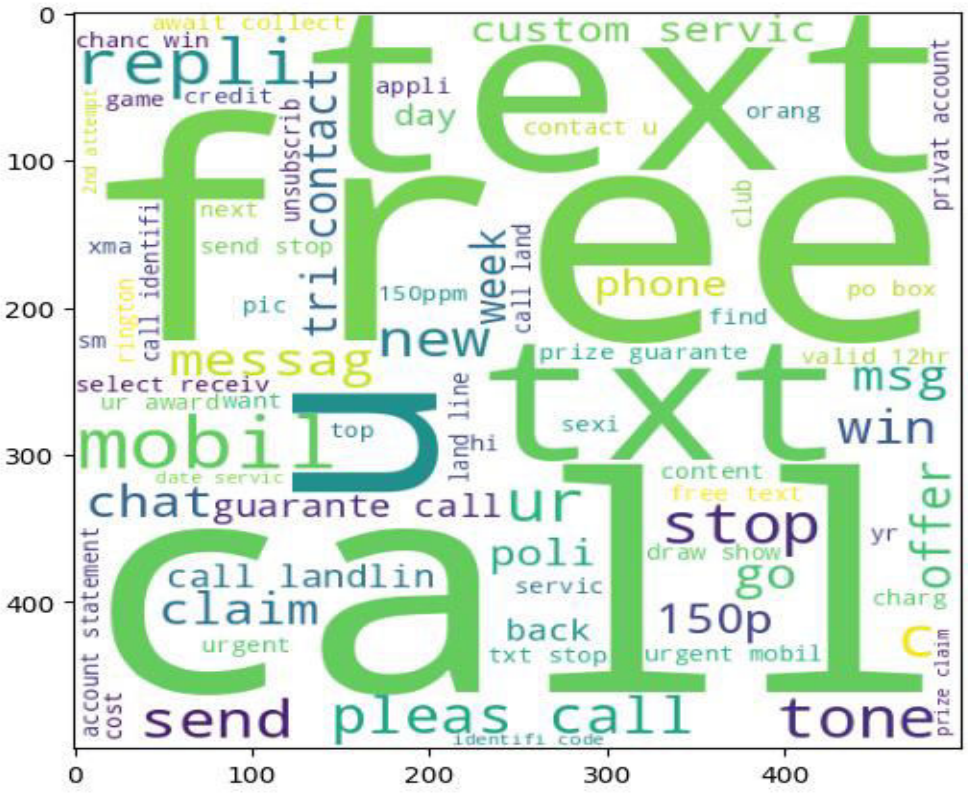
Distribution of Ham vs. Spam Messages



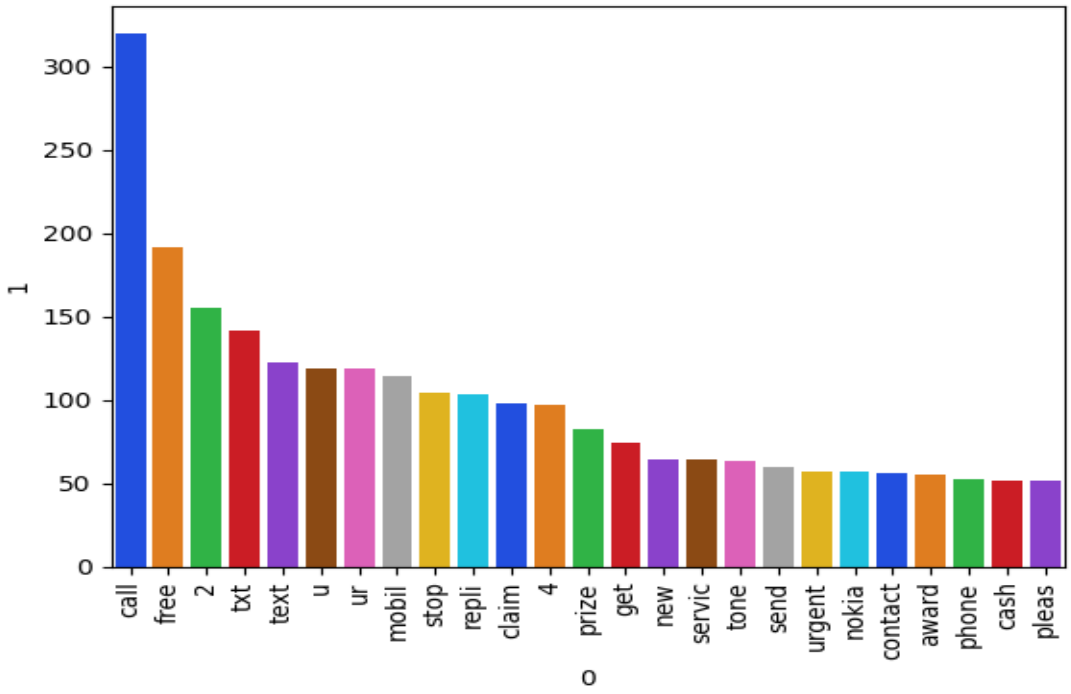
Correlation Matrix of Target and Text Features



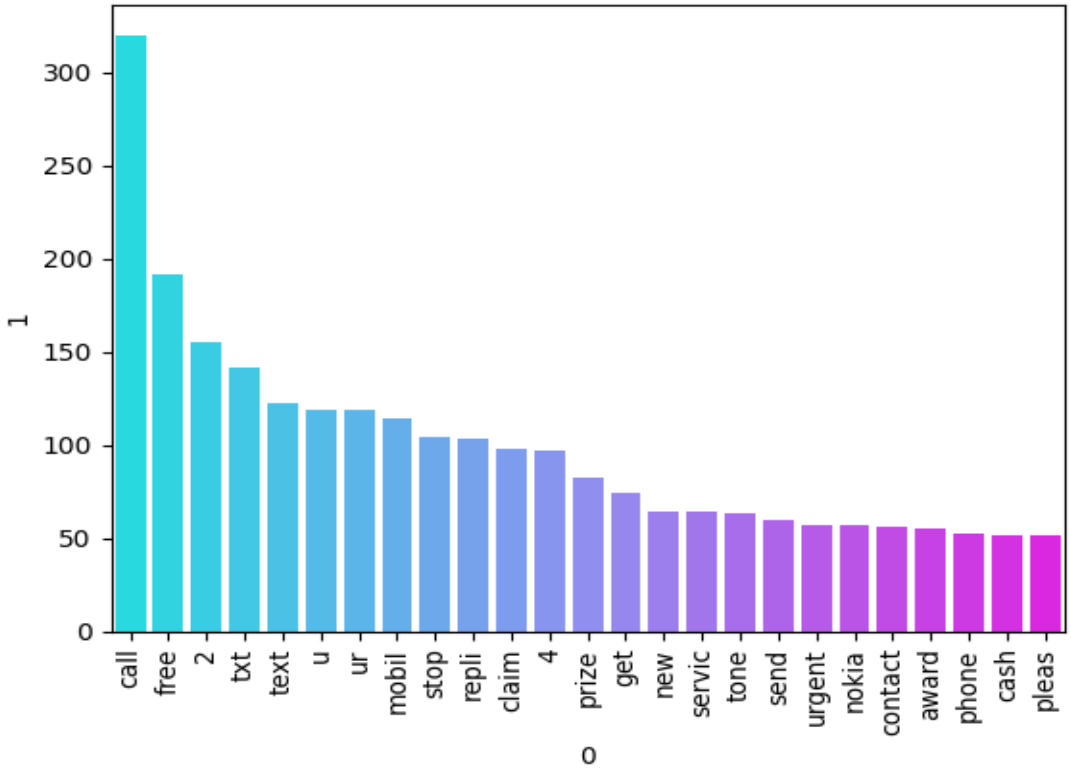
Word Cloud For Spam Messages



Top 25 Most Frequent words in Spam Messages



Top 25 most Frequent Words in Ham Messages



VI. Conclusion

This project successfully developed an SMS spam detection system using a Support Vector Machine (SVM) classifier. A robust NLP preprocessing pipeline was implemented to clean and standardize the text data, and TF-IDF vectorization was used for effective feature extraction. The model was trained and validated on the well-known UCI SMS Spam Collection dataset. The system demonstrated excellent performance, achieving an accuracy of *97.58%* and a precision of *97.48%* , making it a highly reliable tool for distinguishing between spam and ham messages. The exploratory data analysis provided valuable insights into the distinct linguistic patterns of spam and legitimate messages. Furthermore, a practical prediction function was created, showcasing the model's potential for real-world application in enhancing mobile communication security and user experience.

References

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring
COVID-19 through intensive investigation with supervised machine learning algorithm. In

Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.
[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International
Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer
Nature Switzerland.
[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on
colorectal cancer with SMOTE on machine learning Algorithm
[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve
1Professor, Department of computer Science & engineering, Anurag University, TS, India.
2Student, Department of computer Science & engineering, Anurag University, TS, India.
[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time
Object Detection in Drone Surveillance Using YOLOv5," in Proc. 2025 3rd Int. Conf. IoT,
Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6,
doi: 10.1109/ICICAT68430.2025.11414670.
[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis
of Big Data Analytics in Intelligent Transportation Frameworks," in Intersection of Artificial
Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications
in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer,
2025, doi: 10.1007/978-3-031-88304-0_79.
[7] R. D. Kumar, V. N. S.Manaswini, "Applications of blockchain in smart cities: detecting
fake documents from land records using blockchain technology," in Blockchain for Smart
Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.
[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," *International Research
Journal of Modernization in Engineering Technology and Science*, vol. 7, no. 7, July 2025,
doi: 10.56726/IRJMETS81618.
[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system
using LightGBM and decision tree machine learning models," *Journal of Machine and
Computing*, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.
[10] Ravi Kumar Banoth, **Dr. B.V. Ramana Murthy**, "Smart agriculture through IoT and

machine learning for analyzing carbon footprints," in *Proc. Int. Conf. Computer Science and*
Communication Engineering (ICCSCE), Apr. 2025.**[11]** Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer
learning approach: MobileNetV2 with CNN," *SN Computer Science*, vol. 5, art. no. 199,
2024, doi: 10.1007/s42979-023-02500-x.