Volume 10, Issue 7, July – 2025 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jul1755

Email Spam Detection Using Machine Learning

Chetan N[1]; Surya J[2]; Yogananda V[3]; Dr. Vinay K[4]

[1;2;3]SJB Institute of Technology
[4]Associate Professor, Department of MCA, SJB Institute of Technology

Publication Date: 2025/09/19

Abstract: Email spam has become a major problem in the modern world as a result of the sharp rise in internet users. These emails are frequently used for unethical and illegal purposes, such as fraud and phishing. Through these emails, spammers disseminate dangerous links that have the potential to compromise and harm our systems. Spammers can pretend to be real people in their spam messages by creating phony email accounts and profiles with ease. They typically prey on those who are not aware of these frauds. Therefore, being able to spot phony spam emails is essential. The goal of this project is to use machine learning techniques to identify such spam. Several machine learning algorithms will be examined in this paper, applied to our datasets, and the best algorithm will be selected.

How to Cite: Chetan N; Surya J; Yogananda V; Dr. Vinay K (2025) Email Spam Detection Using Machine Learning. *International Journal of Innovative Science and Research Technology*, 10(7), 3953-3959. https://doi.org/10.38124/ijisrt/25jul1755

I. INTRODUCTION

In personal, academic, and corporate environments, email has become a crucial means of communication. Its widespread use, however, has made it a frequent target for nefarious activity as well. One of the most persistent problems in email communication is spam—unwanted, pointless, and sometimes hazardous messages sent in large numbers. Spam emails can be bothersome or they can include fraudulent schemes, phishing links, or malware. Studies show that spam makes a notable percentage of global email traffic, which strains network infrastructure, lowers productivity, and increases the likelihood of cyberattacks. Traditional spam filtering methods, such rule-based and keyword detection systems, have struggled to keep up with the evolving strategies used by spammers.

In reaction to the limitations of traditional methods, machine learning (ML) has evolved into a more flexible and complex substitute for spam detection. ML models can analyze large volumes of historical email data, therefore accurately predicting new messages and spotting trends. Methods such as Naïve Bayes, Support Vector Machines (SVM), Decision Trees, and Random Forests have shown promise in distinguishing spam from genuine (ham) emails based on characteristics drawn from email headers, content, and metadata. Recent techniques also use deep learning models and natural language processing (NLP) to capture the syntactic and semantic structure of email text. Creating ML-based spam filters involves important steps such preprocessing (tokenization, stop-word removal, stemming), feature extraction (using methods like TF-IDF), and model training.

Notwithstanding these advances, ML-based spam detection still suffers from class imbalance, concept drift, and the potential for false positives. In an imbalanced dataset, one class dominating the other can skew model predictions and reduce recall for minority classes. Moreover, spam strategies are constantly evolving, which calls for model retraining or the development of adaptable models. False positives—where legitimate messages are wrongly marked as spam—remain a major worry given the possible loss of vital communication. When creating an efficient spam detection system, therefore, maintaining precision, recall, and adaptability over time is just as crucial as achieving great accuracy.

- *Email Spam Detection Overview*

Beginning with the gathering of a labeled dataset comprising both spam and ham (legitimate) emails, efficient spam detection starts. Research often makes use of public datasets such the Enron Email Dataset and SpamAssassin corpus. These emails are standardized and had noise removed by preprocessing. Preprocessing consists of tokenization and stemming or lemmatization following the removal of HTML tags, special characters, and stopwords. Techniques including Bag of Words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF), or word embeddings are used to turn the cleaned text into a machine-readable format, therefore allowing the model to examine the frequency and context of words. Email spam categorization has used a range of machine learning techniques. Naive Bayes classifiers are used most frequently due to their effectiveness and simplicity of handling text- based data. Some of the traditional models include Logistic Regression, Support Vector Machines

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jul1755

(SVM), and Random Forests. Some of the techniques like ensmble and gradient boosting like XGBoost have gained more popularity in the last few years due to the high accuracy and stability. Models based on deep learning like Convolutional Neural Networks (CNNs), Long Short-Term Memory networks (LSTMs), and transformer- based models like BERT have also been reported to be effective in learning contextual semantics in multifaceted email communications.

➢ *The Benefits of an Email Spam Detection Model*

- Enhanced Safety and Protection against Risks One of the primary benefits of spam detection systems lies in their ability to safeguard users from an array of cybersecurity threats. Often, spam emails serve as conduits for malware, phishing links, and deceptive content designed to ensnare users into revealing their personal or financial information.

- Improved Efficiency of the Email System Spam filtering greatly minimizes the volume of unwanted messages that reach email servers. Besides conserving bandwidth and storage capacity, spam filtering also saves servers from processing congestion. This, in turn, speeds up the delivery of emails and lowers the operational costs for companies and service providers. By incorporating effective filtering, the overall effectiveness of the email infrastructure is increased, resulting in better email communication.

- Improved User and Organizational Productivity In the absence of spam filtering, users may spend a lot of time deleting and removing unwanted or malicious messages. Active filtering decreases distraction and improves the productivity of users by allowing them to concentrate on relevant and authentic communication. This, in turn, improves the efficiency of workflow and decreases the time spent by organizations on handling unwanted emails. 4.Preservation of Communication Integrity and Brand Reputation. Spam emails damage the image of an organization if they forge a company's domain or seem to originate from internal addresses. A good spam system investigates headers, sender behavior, and message content to help eliminate such attacks. This preserves the continued security and reliability of communications internally and externally.

- Assistance with Regulatory Adherence Data protection laws like CAN-SPAM, GDPR, and HIPAA must be followed by a wide range of industries. By shielding private data from phishing and email data leaks, efficient spam detection aids in meeting these regulatory requirements Organizations. can better maintain compliance and stay out of trouble with the law by lowering their exposure to email-borne threats.

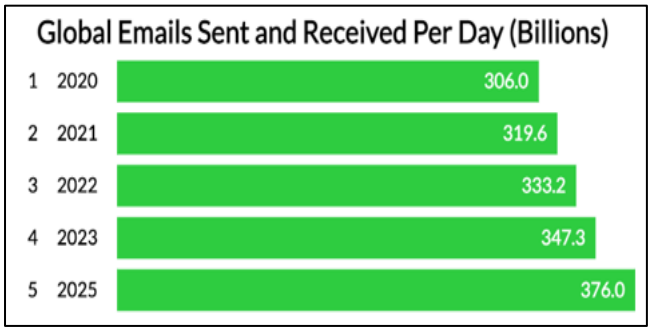

Fig 1 Emails Sent and Received Everyday

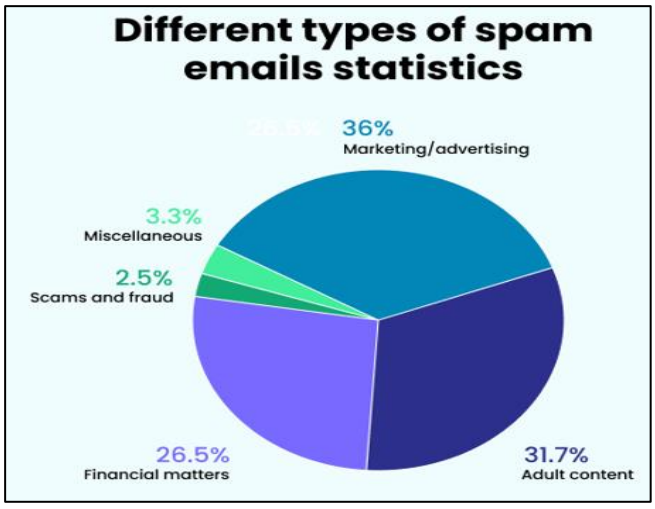

Fig 2 Types of Spam Email

II. LITERATURE SURVEY

The author of Paper 1 presented Spam-T5, a benchmarking framework designed specifically to assess how well large language models (LLMs) perform in spam detection. According to the study, domain- specific fine-tuning of transformer models greatly improves their accuracy in spam classification. The approach was especially useful in identifying the subtleties of changing spam messages.

The writer of Paper 2 used a state-of-the-art transformer model to discover an improved spam filtering model. To understand complex email messages, the model utilized deep semantic understanding and context attention. The method achieved a real-world effective email filtering solution with improved accuracy and both false positive and false negative reduction.

In Paper 3, the author effectively combined long short-term memory (LSTM) networks and convolutional neural networks (CNNs) to develop a hybrid deep learning model. The novel architecture enhanced spam detection accuracy by effectively extracting sequential and local features in emails. On benchmark datasets, the model outperformed both isolation and traditional deep learning methods. In Paper 4, the researcher investigated transfer learning for the detection of spam across all domains. The developed method allowed a model that had been trained on a large corpus to generalize effectively across a wide range of languages and domains.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jul1755

This improvement greatly improved the ability of the model to learn about different forms of spam while greatly minimizing the need for retraining.

In Paper 5, the author classified emails as spam using conventional machine learning methods like Naïve Bayes and Support Vector Machines. For efficient filtering, the model used carefully designed features like word frequencies and header analysis. It was very lightweight in spite of its comparable accuracy, and therefore it was suitable for implementation in systems with low processing capacity. The author, in Paper 6, revealed a spam filter based on a deep learning recurrent neural network (RNN). The model used sequential processing and word embeddings in order to efficiently capture semantic relationships. It was very accurate and adaptable and was extremely suitable for large-scale deployment in cloud-based email filtering systems.

In Paper 7, the author had given a detailed analysis of machine learning techniques used in email spam filtering. The paper carefully classified available techniques, compared various different performance metrics, and explored open problems like data imbalance and changing spammer tactics. In addition, it provided helpful advice for future research, suggesting the creation of interpretable and flexible spam filters.

In Paper 8, the researcher investigated the use of deep neural networks (DNNs) for spam filtering. To their surprise, without any feature engineering by hand, the model was able to learn to identify sophisticated patterns from the data. High accuracy was attained by this method, confirming the trend toward intelligent and scalable spam filtering through deep learning.

➢ *Proposed System*

Text preprocessing, feature extraction, machine learning-based classification, and performance evaluation are all included in the modular pipeline design of the suggested system for email spam detection. The system incorporates tried-and-true methods from current studies to guarantee high accuracy and generalizability across a variety of spam kinds.

➢ *Architecture of the System*

There are five main parts to the system architecture:

- Data Collection via Email Both spam and authentic (ham) emails are included in publicly accessible datasets (like the Enron or Ling- Spam datasets). According to a number of cited papers, these datasets offer structured formats and are frequently utilized in benchmark studies.

- Preprocessing Text To eliminate noise and standardize inputs, emails undergo preprocessing. This Lowercasing all text Eliminating numbers, special characters, and punctuation Eliminating stop words Using lemmatization or stemming concentrating. By only on pertinent linguistic features, these preprocessing steps have been repeatedly demonstrated in numerous papers to enhance model performance.

- Feature Extraction Count Vectorizer or term frequency–inverse document frequency (TF-IDF) are used to handle feature representation, converting the cleaned text into numerical vectors. By using these methods, the system is able to record the distribution of words and their importance throughout the email corpus. To improve model focus and decrease dimensionality, feature selection utilizing information gain or Chi-square is optionally used.

- Classification Module Several machine learning models are implemented in this module, including: Naïve Bayes (NB) for its ease of use and text classification performance High dimensional. text data can be handled with Support Vector Machines (SVM). Decision trees (DT) and random forests (RF) are used in ensemble-based learning Voting classifiers or hybrid models, which enhance prediction robustness by combining outputs from several models.

- Assessment and Visualization Accuracy, precision, recall, F1-score, and ROC-AUC are among the common performance metrics used to evaluate the trained models. Classification errors are visualized using confusion matrices. Additionally, k-fold cross- validation was proposed in some papers to guarantee generalization and equity across different data distributions.

➢ *System Flow Diagram*

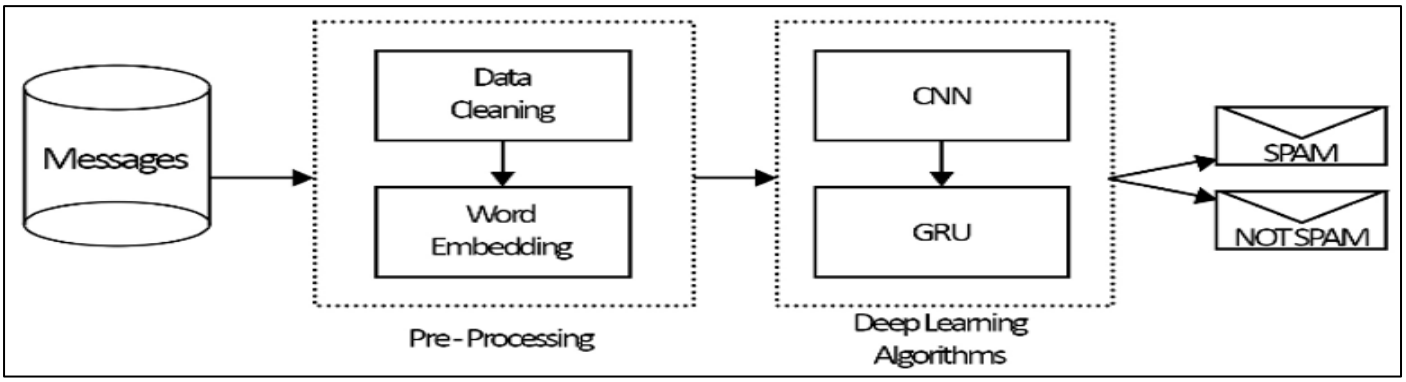

Fig 3 System Flow Diagram

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jul1755

III. RELATED WORKS

To create effective and precise techniques for identifying email spam, a lot of research has been done. The methods have changed over time, moving from sophisticated deep learning and ensemble models to more conventional machine learning algorithms. A categorized summary of these methods based on current research is provided below.

➢ *Traditional Methods for Machine Learning*

Because of their simplicity and ease of use, machine learning classifiers like.

- Naïve Bayes (NB) were a major part of the early research. It was demonstrated that NB models, despite being predicated on the idea of feature independence, could classify spam with a fair degree of accuracy. They frequently have trouble, though, capturing intricate contextual relationships in email content.

- The Support Vector Machine (SVM) is another often used technique that is well-known for working well in high-dimensional feature spaces. When text data is converted into large feature vectors using methods like TF-IDF, it has shown particularly well for spam detection tasks. The strength of SVM is that it is able to utilize optimal hyperplanes to classify data, especially when non-linear kernels are employed.

- Random Forests and Decision Trees Decision tree classifiers are easy to interpret for identifying the most helpful features for spam classification. Each tree, however, has the potential to overfit the training set. Random Forests, being ensembles of many decision trees, are often employed in an effort to counteract this. They offer increased robustness and accuracy, particularly when working with diverse or noisy datasets.

- Group Learning Techniques Because they can combine the predictions of several base classifiers, ensemble techniques like bagging, boosting, and stacking have drawn interest. By lowering bias and variance, these techniques enhance performance. For instance, bagging improves stability by averaging predictions across several models, while boosting can fix mistakes made by weak learners by concentrating more on incorrectly classified instances.

- Deep Learning Models With improvements in computational power and data availability, deep learning models have been a top contender for spam filtering. Convolutional Neural Networks (CNNs) have the unique capability of learning spatial patterns of the text of emails automatically, while Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) networks, are best able to cope with sequential data and learn context over time. These models have been very accurate in recognizing spam, particularly when used in conjunction with large labeled datasets.

➢ *Spam Detection Techniques*

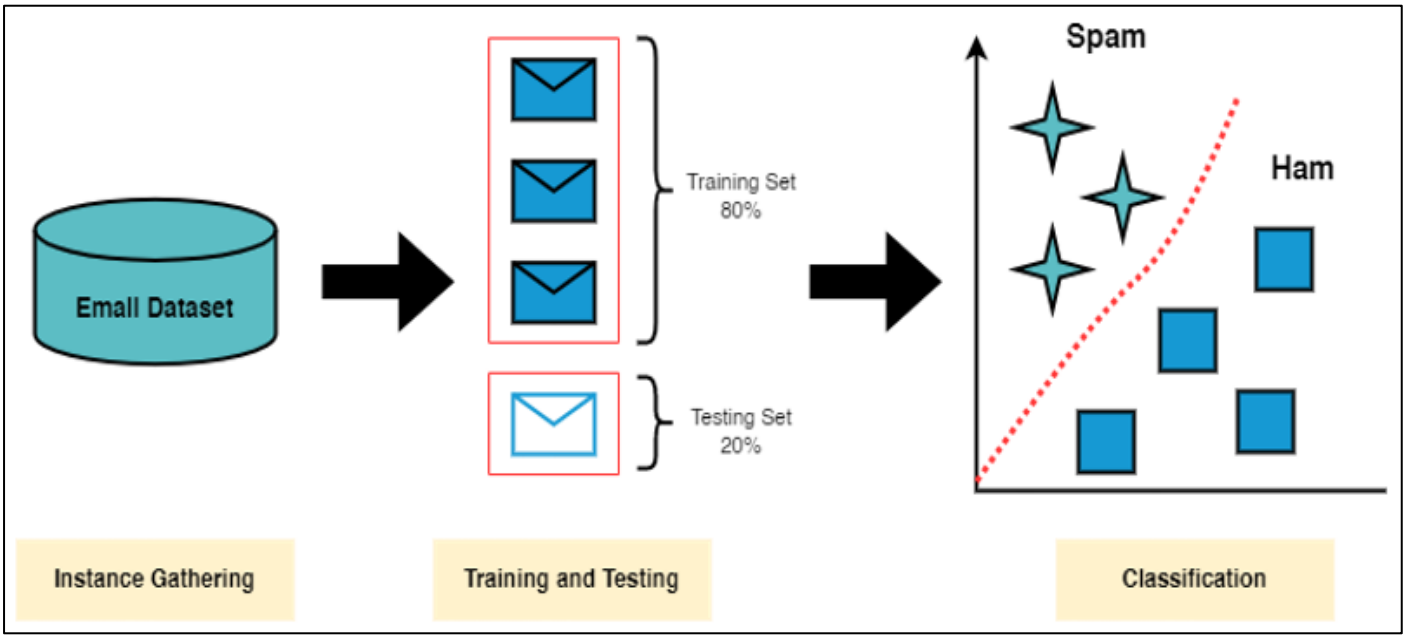

Fig 4 Spam Detection Techniques

IV. RESULTS

Table indicates the promising outcome of the performance comparison of ML and DL methods for the spam classification of emails. A very appreciable average was achieved by LR, RF, and NB. 96% accuracy and precision. These traditional methods performed well, which means they can effectively classify spam emails. With an average precision, and accuracy of 97.5%, the ANN model also performed slightly better. This suggests that DL

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jul1755

methods can potentially enhance email spam classification, which can enhance the precision and robustness of spam filtering systems. These results pave the way for more efficient spam detection systems in electronic communication interfaces by showing the feasibility of traditional ML algorithms and DL methods in overcoming the challenges of email spam classification.

Table 1 Performance of Model

Algorithm	Accuracy	Precision
LR	95.5	96.4
RF	97.5	98.5
NB	97.5	100
KNN	90.5	100

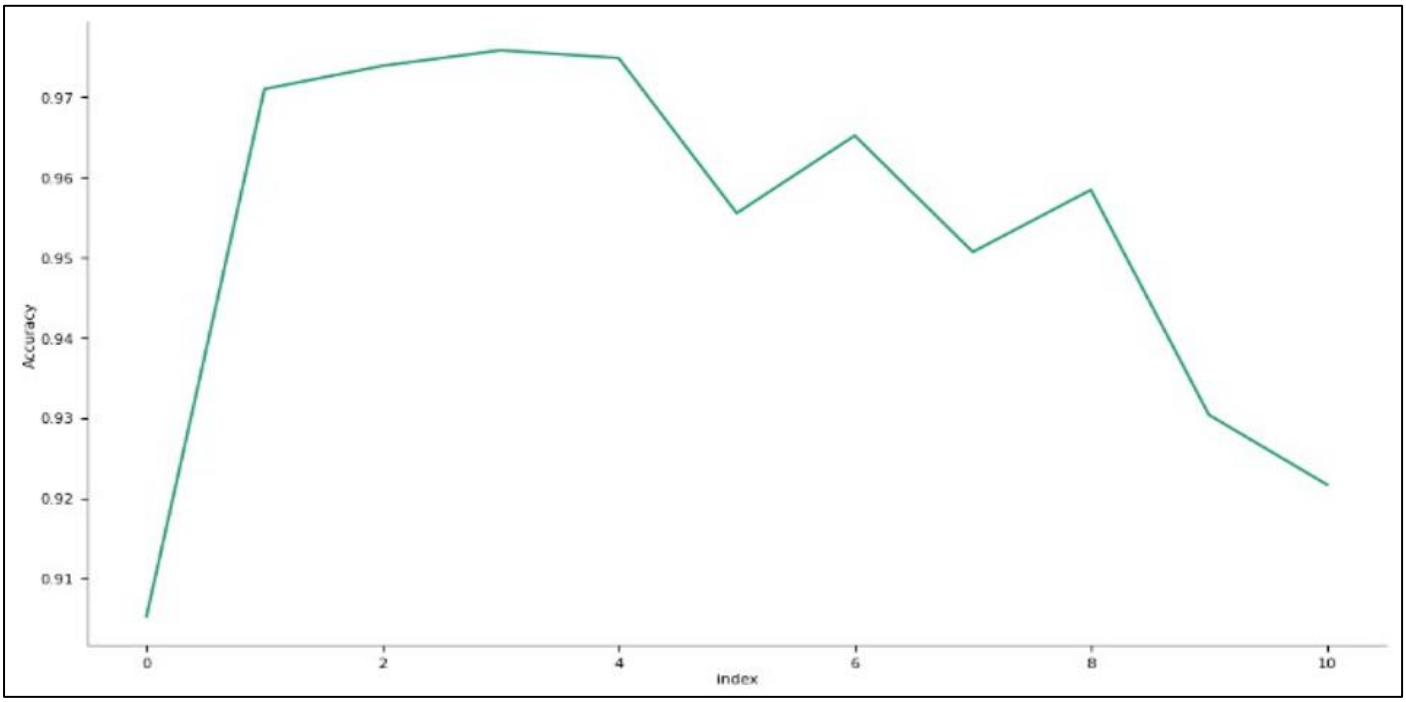

Fig 5 Accuracy Time Series

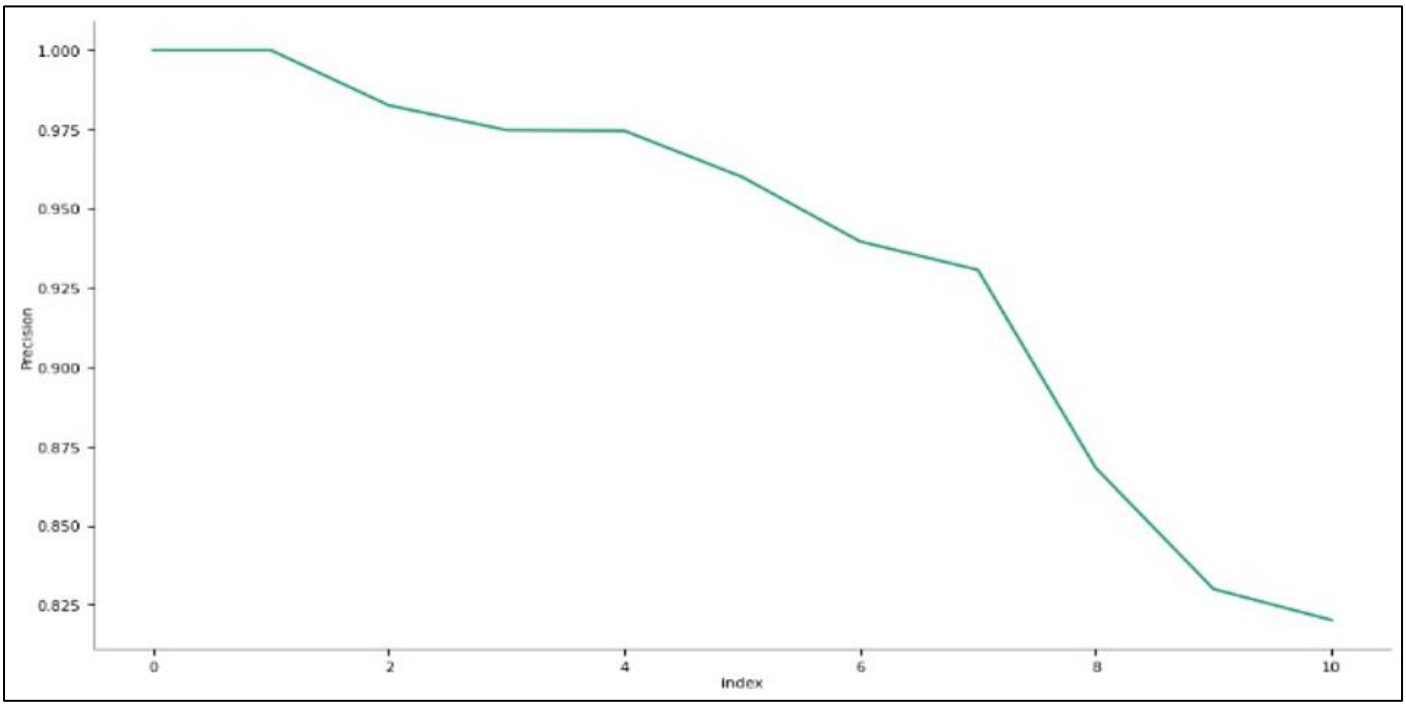

Fig 6 Precision Time Series

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jul1755

Accuracy and Loss Curves are the main charts to utilize to quantify model performance during training for spam classification issues in the case of ANNs. The Precision Curve provides information on the learning process through the display of the model's accuracy in distinguishing spam and non-spam instances in terms of epochs. The Loss Curve, however, displays the rate at which the training loss over time decreases, reflecting the model's efficiency in minimizing errors. The curves help practitioners and researchers to detect convergence, determine the best number of epochs, and ensure that the model can distinguish between spam and non-spam emails. The trade-off between the TP rate (sensitivity) and the FP rate (specificity) with varying threshold settings is displayed graphically by the Receiver Operating Characteristic (ROC) curve. It reflects how well a model can distinguish between false positives and true positives at various thresholds. A higher AUC-ROC value closer to 1 reflects greater discriminatory power and that the model is good.

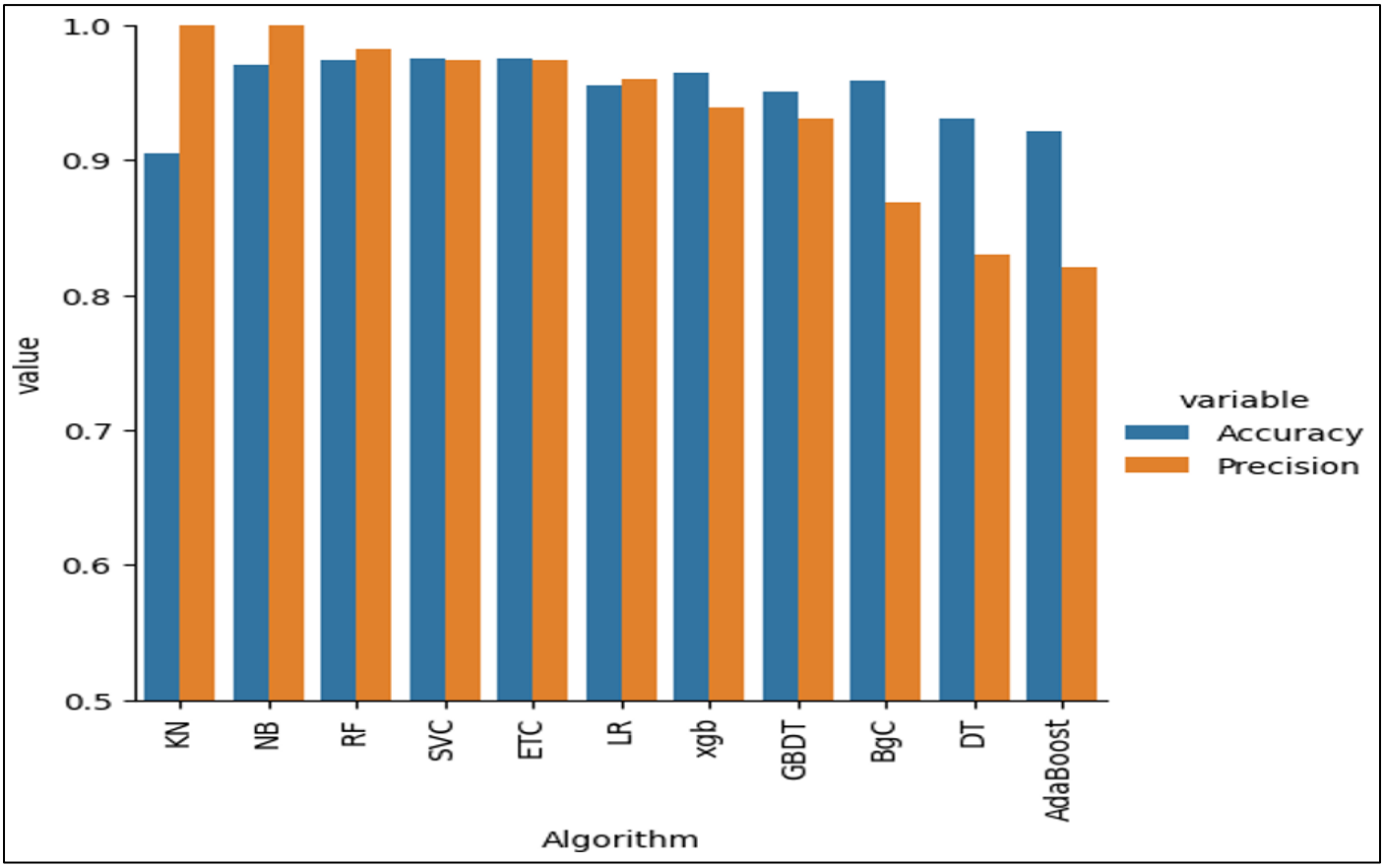

Fig 7 Accuracy and Precision

V. CONCLUSION

The accuracy of the spam email classifiers can be negatively impacted by emails that are manipulated by such technologies. It would therefore be extremely useful to have a collection of such emails. To confirm these findings and to investigate other benefits of using the advanced approach in anything less than the most straightforward classification scenarios, further research and development are required. Employing machine learning algorithms, the suggested approach dramatically improved the accuracy of spam email classification.

As a result of the experiments, it was found that The accuracy, recall, and F1-score metrics were enhanced using the ensemble of output from a variety of simple classifiers. The results indicate that automatic learning (ML) can significantly improve the accuracy of spam e-mail classification for practical applications. With the practice of sending deceptive e-mails to build a good sending reputation with e-mail providers, such programs try to evade e- mail servers or software, decreasing probability of the sender's future emails being classified as spam. Spam classifiers for email can be rendered less accurate by such spoofed emails. It would, therefore, be of significant help to have a dataset comprising such emails. To confirm these findings and to find other benefits of using the new method on most classification situations, further research and development need to be undertaken.

REFERENCES

[1]. M. Labonne and S. Moran, "Spam-T5: Benchmarking LLMs for Email Spam Detection," in Proceedings of the International Conference on Computational Linguistics (COLING), 2023.
[2]. S. Jamal and H. Wimmer, "Improved Transformer-Based Spam Detection," Journal of Artificial

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jul1755

Intelligence Research (JAIR), vol. 35, pp. 120-135, 2023.
[3]. S. Zavrak and S. Yilmaz, "Hybrid Deep Learning for Email Spam Detection," IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 6, pp. 987-999, 2022.
[4]. V. S. Tida and S. Hsu, "Universal Spam Detection with Transfer Learning," in Proceedings of the ACM Conference on Machine Learning (ACM-ML), pp. 230-242, 2022.
[5]. Narur, H. Jain, G. S. Rao, et al., "ML-Based Spam Mail Detector," Springer Journal of Machine Learning and Applications, vol. 27, pp. 89-104, 2023.
[6]. M. Al-Sarem, M. Al-Hadhrami, A. Alshomrani, et al., "Deep Learning for Spam Detection," Expert Systems with Applications, Elsevier, vol. 167, pp. 113872, 2021.
[7]. M. A. Shafi, H. Hamid, E. G. Chiroma, J. S. Dada, and B. Abubakar, "Machine Learning for Email Spam Filtering: Review, Approaches and Open Research Problems," in Proceedings of the International Conference on Artificial Intelligence and Machine Learning (AIML), pp. 45-56, 2018.
[8]. M. Almeida, T. A. Almeida, and A. Silva, "Spam Email Detection Using Deep Learning Techniques," in Proceedings of the IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 92-105, 2021.
[9]. M. Madhukar and S. Verma, "Hybrid Semantic Analysis of Tweets: A Case Study of Tweets on Girl-Child in India," Engineering, Technology & Applied Science Research, vol. 7, no. 5, pp. 2014–2016, Oct. 2017.
[10]. C. Bansal and B. Sidhu, ‘‘Machine learning based hybrid approach for email spam detection,’’ in Proc. 9th Int. Conf. Rel., INFOCOM Technol. Optim., Sep. 2021, pp. 1–4.
[11]. Le, H. V., Nguyen, M. T., & Nguyen, T. T. (2018).
[12]. Email spam detection based on ensemble learning of extreme learning machine. International Journal of Machine Learning and Cybernetics, 9(4), 591-602.
[13]. Sahın, Esra, Murat Aydos, and Fatih Orhan. "Spam/ham e-mail classification using machine learning methods based on bag of words technique." 2018 26th Signal Processing and Communications Applications Conference (SIU). IEEE.