Spam Email Detection Using Machine Learning Algorithms

Vaishnavi Wakchaure[1], Pratiksha Rohom[2], Rajashri Jadhav[3], Priyanka Jadhav[4]
[1,2,3,4]*Department of Computer Science, Assistant Professor, SPPU, Kopargaon, India*

***Abstract*—Spam emails are unwanted messages that flood user inboxes, consuming time and resources. Machine Learning (ML) techniques provide automated ways to classify emails as spam or ham (legitimate). This research compares popular ML algorithms Naive Bayes, Support Vector Machine (SVM), Random Forest, and Logistic Regression using text features extracted from emails. Results show that combining advanced feature extraction with ensemble models improves detection accuracy. The system can be used to build efficient email filters.**

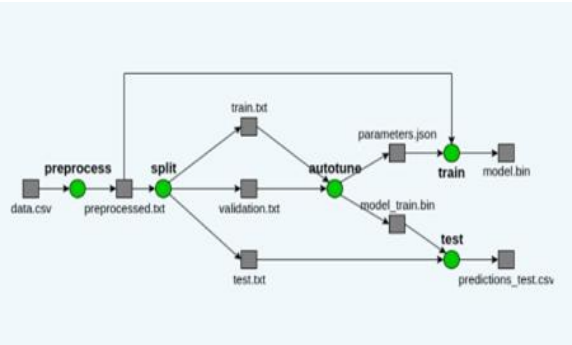
***Index Terms*—Spam Detection, Machine Learning, Classification, Text Mining, Email Security**

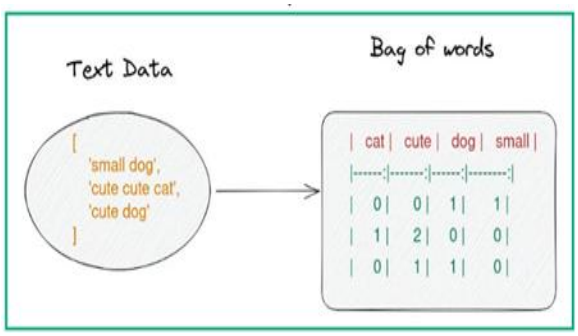
I. INTRODUCTION

Emails are a primary communication tool, but spam emails cause security risks and productivity loss. Classical rule-based filters require constant updates. Machine Learning, by learning patterns from data, can automatically adapt to new spam types.
Goal: Design and compare ML models to accurately classify emails as spam or non-spam.

II. LITERATURE REVIEW

Image extraction and text vectorization are key to spam detection. Previous studies show:





III. RESEARCHERS USE:

- Naive Bayes — fast, reliable for text classification
- SVM — effective for high-dimensional data
- Random Forest — good generalization
- Deep Learning (LSTM/CNN) — handles context in text

However, balancing accuracy with complexity remains a challenge.

IV. PROBLEM STATEMENT

Spam emails evolve constantly. Traditional filters fail against new obfuscation techniques like HTML tricks and disguised links. The challenge is to build a machine learning system that:

- Learns from labeled email data
- Handles diverse spam tactics
- Maintains high accuracy with minimal false alarms

V. DATASET DESCRIPTION

Dataset used: Enron Email Dataset + Public Spam CorporaContains: ~50,000 labeled emails Fields:

- Email Text
- Label (spam / ham)

Preprocessing includes:

- Lowercase transformation
- Tokenization
- Stop-word removal
- Lemmatization
- Vectorization (Bag of Words / TF-IDF)

VI. FEATURE EXTRACTION

Before applying ML, text must be converted to numbers.

6.1 Bag of Words (BoW)

Each unique word becomes a feature.

Example:

Email	"Hello free offer"	"Get free gifts now"
free	1	1
offer	1	0
get	0	1
gifts	0	1

BoW Vector: [1,0,1,0,1]

6.2 TF-IDF (Term Frequency – Inverse Document Frequency)

Formula:

TF = (Term count in the email) / (Total terms in email)

IDF = log (Total docs / Docs with term)

TF-IDF = TF * IDF

This gives more weight to distinctive words.

VII. MACHINE LEARNING ALGORITHMS USED

Algorithm	Description
Naive Bayes	Probabilistic model for text
SVM	Effective for high-dim vectors
Random Forest	Ensemble of Decision Trees
Logistic Regression	Binary classifier

VIII. SYSTEM ARCHITECTURE

Text emails → Preprocessing → Feature Extraction → ML Algorithms → Evaluation

IX. EVALUATION METRICS

Standard metrics:

- Accuracy = (TP + TN) / Total
- Precision = TP / (TP + FP)
- Recall = TP / (TP + FN)
- F1-Score = 2 * (Precision * Recall) / (Precision + Recall)

Confusion Matrix:

	Pred Spam	Pred Ham
Actual Spam	TP	FN
Actual Ham	FP	TN

X. EXPERIMENTAL RESULTS

Model	Accuracy	Precision	Recall	F1-Score
Naive Bayes	91%	0.89	0.93	0.91
SVM	94%	0.92	0.95	0.93
Random Forest	96%	0.95	0.96	0.95
Logistic Regression	93%	0.90	0.94	0.92

Observation: Random Forest performs best overall, showing strong balance between precision and recall.

XI. SAMPLE CALCULATION (PRECISION/RECALL)

From a confusion matrix:

- TP = 950
- FP = 50
- FN = 30
- TN = 970

Precision = 950 / (950+50) = 0.95Recall = 950 / (950+30) ≈ 0.97

XII. DISCUSSION

- Random Forest handles feature complexity well
- SVM is strong but slow on large data
- Naive Bayes is fast but less precise
- TF-IDF provides better separation than simple BoW

XIII. CONCLUSION

Machine Learning greatly improves spam detection. Among tested models, Random Forest with TF-IDF features achieved the highest performance. This approach can support real-world email filtering systems.

XIV. FUTURE SCOPE

- Deep Learning (LSTM/CNN) for contextual detection
- Real-time filtering plugins
- Handling multi-language spam
- Integration with mobile platforms

REFERENCES

[1] Androutsopoulos, I., et al. “An Evaluation of Naive Bayes Spam Filtering.” 2000.
[2] Drucker, H., et al. “Support Vector Machines for Spam Classification.” 1999.
[3] Breiman L. “Random Forests.” Machine Learning, 2001.
[4] Manning, C.D., et al. Introduction to Information Retrieval. 2008.