WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Spam Detection using Machine Learning

***Drishti Lalwani*[1], *Ishani Pandey*[2], *Krupi Saraf*[3], *Harshit Rathore*[4], *Gaurav Tiwari*[5]**

[12345]Computer Science Department Acropolis Institute of Technology and Research Indore,India
drishtilalwani1432@gmail.com[1], pandeyishani84@gmail.com[2], krupisaraf@acropolis.in[3], harshitrathore20384@acropolis.in[4], gauravtiwari20162@acropolis.in[5]

ABSTRACT—

Developing a robust spam detection system that involves accurately identifying and filtering spam messages, minimizing false positives, and adapting to the dynamic strategies used by spammers using various machine learning algorithms. Spam is becoming more and more common, which puts users' security, privacy, and overall experience at risk. This is due to the exponential rise of online communication channels like social media, messaging and email. Spam can take different forms, but it is typically defined as uninvited, irrelevant, or malicious content. This model can discriminate between spam and non-spam communications with high precision by continuously learning from labelled datasets and making adjustments for providing accurate results.

Keywords—Preprocessing, Machine learning, Multinomial Naive Bayes, Support Vector Classifier, ExtraTrees Classifier.

INTRODUCTION

In today's digital age, our inboxes are flooded with messages. Unwanted and bothersome communications that flood our social media accounts, messaging applications, and email accounts are known as spam, and they can be dangerous in addition to being distracting. Envision being protected from spam by a digital guardian that filters through messages and marks the ones that are spam. In fact, a spam detection system performs precisely. To ensure that only the communications user need and want arrive in their inbox, it functions as a kind of intelligent filter that sorts through incoming messages. Not only spam be disturbing, but it also carries significant risks. There are legitimate risks associated with spam, ranging from phishing efforts to damage user device with malware-filled files to personal information theft. A spam detector is a first line of defense against these threats, a spam detector gives security and helps in avoid wasting time or aggravation. Spam comes in many forms, each with its own characteristics and dangers. Email Spam are the most common type, email spam inundates your inbox with unsolicited messages advertising products, services, or fraudulent schemes. Social Media Spam are on social media platforms, spam can take the form of fake accounts, automated bots, or spammy comments and messages promoting links or products. Text Message Spam also known as SMS spam, this type targets user's phone with unwanted text messages, often containing phishing attempts or fraudulent offers. Messaging apps aren't immune to spam either. Spam messages can flood your chat windows, trying to lure you into clicking malicious links or sharing personal information. Continued advancements in artificial intelligence and machine learning have led to even more sophisticated Spam Detection Systems capable of adapting to new spamming techniques in real-time. This project is easy to use and less complex. It classifies spam and non-spam effectively. This simplicity enhances user experience and make the system more accessible to a wider range of users. The model has been built using machine learning algorithms which helps the model to give results with accuracy. It also provides results with high precision while maintaining ease of use. It is a useful tool in the fight against spam as it provides a good mix between usability and speed.

LITERATURE REVIEW

This Paper [1] establishes the basis for statistical techniques in spam detection by introducing a Bayesian approach to email spam filtering. The authors suggest a probabilistic model that determines the probability that an email is spam. The outcomes of the experiments show how well the Bayesian classifier performs in reliably identifying spam from authentic messages.

This paper focuses on using Deep Learning for Email Spam Detection [2] explores the application of deep learning techniques, specifically convolutional neural networks (CNNs), for email spam detection. The authors propose a CNN-based model that learns discriminative features from email content to classify messages as spam or non-spam. Experimental evaluations demonstrate the superiority of the deep learning approach in handling complex spam patterns compared to traditional methods.

In this study [3], the authors present a hybrid spam detection system that combines multiple classifiers to improve detection accuracy. The hybrid model integrates rule-based filters with machine learning algorithms, such as Naive Bayes and support vector machines (SVMs), to effectively identify spam emails.

This paper [4] explores the phenomenon of adversarial spam, characterized by sophisticated evasion techniques and targeted attacks on social media platforms. The authors studied how these spammers use tricky tactics to avoid getting caught and to make people engage with their spammy content. By looking at this from different angles, they showed how tough it is to stop this kind of spam. They also stressed how important it is to come up with strong ways to fight against it.

This research [5] focuses on improving email spam filtering with lightweight user interaction and reinforcement learning

and talks about a new way to stop spam emails from reaching user inbox. They suggest a system that works like this: when user mark an email as spam or not spam, the system learns from your choices and gets better at spotting spam in the future. They tested this idea and found that it works well. So, basically, by learning from what you do, the system becomes smarter at catching spam, making it easier for user to manage your emails and be happier with their email experience.

METHODOLOGY

Data Gathering and Preprocessing: In this step the compilation of a dataset with data such as emails, text etc. are samples of spam and non-spam. The dataset used in our project is obtained from Kaggle [9]. The gathered data is pre-processed using the following steps:

- All the special characters are removed.
- Stop words are removed.
- Porter's Stemming Algorithm is applied to bring the word in their most basic form.
- The word frequency of all the words.
- The normalized term frequency of all the words.
- The inverse document frequency of all the words.
- Term Document Frequency inverse document frequency (TF-IDF): It is the process of transforming textual data into numerical characteristics that machine learning algorithms use.

Feature Extraction: A feature in pattern recognition and machine learning can be defined as an individual characteristic or property of a phenomenon being observed

It is the process of converting textual data into some specific format that can make the important text available for analysis. There are several feature extraction techniques for text classification e.g. bag-of-words, n-grams, TF-DIF etc. In this project we have used Machine Learning algorithms for feature extraction.

Machine Learning techniques have been successful for spam classification. These techniques extract information from labeled training datasets and use this information to train the classifier. Spam email classification is considered as a binary classification problem in which emails are classified as spam or ham. The machine learning algorithms that are most popular in spam classification are Naïve Bayes [13], Support Vector machine, Decision Tree and Neural network.

Model Training and Testing: The machine learning models are trained on the preprocessed dataset using the extracted features. The system is tested on models such as Support Vector Classifier, K-Nearest Neighbors, Naive Bayes (Multinomial Naive Bayes in this context), Decision Tree, Logistic Regression, Random Forest, Adaptive Boosting, Bagging Classifier, Extra Trees Classifier, Gradient Boosting Decision Trees, Extreme Gradient Boosting. There is utilization of the provided accuracy and precision scores of each model to guide the training process.

Model Integration: The model is developed using an integration strategy. Among all the tested models Support Vector Classifier, Multinomial Naive Bayes and Extra tree Classifier have the most accuracy. These three models are integrated together to give the most accurate results. The integration is done using voting classifier model where each model gets a vote, and the final prediction is based on majority voting.

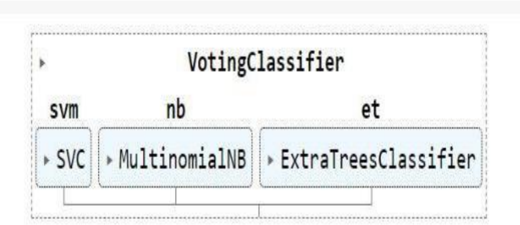

Figure 1.1: Voting Classifier

Model Evaluation: The evaluation of the integrated model is done on a separate validation dataset using appropriate metrics such as accuracy, precision. To test the trained machine, a different CSV file is developed with unseen data i.e. data which is not used for the training of the machine; named emails.csv.

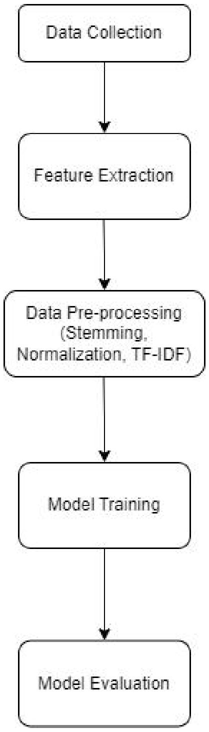

Figure1.2: Flow Diagram

PROBLEM FORMULATION

The constant barrage of spam messages is one of the major challenges that the digital age. Spam interrupts user experience and poses serious security issues, as it can infiltrate social media feeds and clog email inboxes.

The creation of reliable spam detection technologies has the potential to completely transform the way to handle this challenge. The key to overcoming this difficulty is developing spam detection systems that can reliably distinguish between spam and legitimate communications, protecting user privacy and system's integrity. The existing spam detection systems struggle to effectively differentiate between genuine and spam messages, leaving users vulnerable to scams and privacy breaches. The three-layer approach for spam detection:

Gathering Samples: Collect a variety of emails, including both that are consider spam (like advertisements, phishing attempts, or unsolicited emails) and those that are consider legitimate (such as emails from friends, family, or work). Gather a diverse set of spam and non-spam emails to give the system a wide range of examples to learn from.

Identifying Patterns: Analysis these emails closely. Finding common traits or patterns that tend to occur more frequently in spam emails than in non-spam emails. These patterns could include specific words or phrases, unusual sender addresses, certain types of attachments, or formatting characteristics.

Labeling and Categorizing: Based on the patterns, categorize each email as either spam or non-spam. For example, if an email contains a lot of words related to money-making schemes and comes from an unfamiliar sender, label it as spam. Similarly, if an email is from a known contact and contains typical communication content, label it as non-spam. This process helps create a labeled dataset that the system can use to learn the difference between spam and non-spam messages.

RESULT DISCUSSION

The test results for the various models like KNN, random forest, adaptive boosting etc. are

Models	**Accuracy**	**Precision**
Support Vector Classifier	**0.975**	**0.974**
K-Nearest Neighbors:	**0.905**	**1.0**
Multinomial Naive Bayes	**0.970**	**1.0**
Decision Tree	**0.930**	**0.817**
Logistic Regression	**0.958**	**0.970**
Random Forest	**0.975**	**0.982**
Adaptive Boosting	**0.960**	**0.929**
Bagging Classifier	**0.958**	**0.868**
Extra Tree Classifier	**0.974**	**0.9745**
Extreme Gradient Boosting	**0.967**	**0.926**

Table 1.1: Accuracy and Precision of various Models

Building reliable and accurate systems frequently involves integrating machine learning models, particularly in situations like spam detection when a variety of patterns and factors influence the message classification.

In this instance, the Support Vector Classifier, Extra Trees Classifier, and Multinomial Naive Bayes are three of the top-performing models combined in the suggested system. The integrated system achieves an outstanding 98% accuracy and 99% precision by utilizing the capabilities of each model and their capacity to collect different components of the data.

Support Vector Classifier (SVC) is known for its effectiveness in separating classes by finding the hyperplane that maximizes the margin between them. It performs exceptionally well in cases where the data is linearly separable or can be transformed into a higher-dimensional space. The high precision (97.4%) and accuracy (97.5%) of SVC show how reliable it is at distinguishing between emails that are spam and those that are not.

Multinomial Naive Bayes is a probabilistic classifier based on Bayes' theorem. Despite its simplicity, it often performs well in text classification tasks, making it a suitable choice for analyzing email content. With a precision of 100% and accuracy of 97%, it appears to be able to capture the probabilistic correlations between terms and their frequency in spam emails.

Extra Trees Classifier, a variant of Random Forest. It is an ensemble learning method that constructs multiple decision trees and combines their predictions through voting. It excels in handling high-dimensional data and noisy features while reducing overfitting. Extra Trees Classifier has a 97.4% accuracy rate and a 97.45% precision rate, indicating its strong capacity to make accurate predictions and generalize it to new data.
By combining these models, the advantages of each model are enhanced while the disadvantages of each are reduced. The combined decisions made by the Multinomial Naive Bayes, Support Vector Classifier, and Extra Trees Classifier allow the integrated system to outperform each model separately in terms of accuracy and precision. The following are some ways that the integration process leads to better performance:

Diverse Model Architectures: Each model in the integration brings a unique perspective to the classification task due to its underlying algorithms and assumptions. While Multinomial Naive Bayes makes advantage of probabilistic correlations, Support Vector Classifier concentrates on maximizing the margin and Extra Trees Classifier uses ensemble learning techniques. An increased variety of patterns and features in the data can be captured by the integrated system through the combination of various disparate points of view.

Variability: Individual models may perform well on certain subsets of the data but struggle with others. By integrating multiple models, the system becomes more robust to variability in the dataset. If one model fails to accurately classify a particular instance, the decision of other models can compensate for it, leading to more consistent and reliable predictions overall.

Decision Making: The integrated system makes predictions based on the collective decision of multiple models, often through techniques like voting or averaging. By using this strategy, the likelihood of overfitting in data is decreased and the system's capacity to generalize to new, unobserved examples is improved. The integrated system gains from the combined knowledge and experience of Support Vector Classifier, Multinomial Naive Bayes, and Extra Trees Classifier by combining their predictions.

Error Correction and Improvement: In cases where individual models make incorrect predictions, the integrated system can potentially correct these errors by considering the decisions of other models. Through continuous learning and refinement, the system can adapt to evolving patterns in spam emails and improve its performance over time.

Using the advantages of each model to gain more accuracy and precision, the integration of Support Vector Classifier, Multinomial Naive Bayes, and Extra Trees Classifier into a single system provides an excellent solution for spam identification. The integrated system offers an efficient defense against spam emails by integrating a variety of viewpoints, resilience to variability, consensus decision-making and error-correction. This helps users keep their inboxes clear of clutter and safe from harmful threats.

CONCLUSION

To protect users from unsolicited and sometimes hazardous messages, a strong spam detection system is necessary. The proposed system can efficiently filter spam emails, texts, and other types of unwelcome content by using sophisticated algorithms and machine learning approaches. The system may change adapt new spamming strategies by continuously observing and analyzing patterns, phrases, and sender behavior. Furthermore, user feedback and human verification procedures can improve spam detection accuracy even more.

The proposed spam detection system that combines the Multinomial Naive Bayes, Extra Trees Classifier, and Support Vector Classifier (SVC) is a strong barrier against unsolicited email clutter. Utilizing the distinct advantages of every model and synthesizing their combined expertise, the system attains a remarkable accuracy and precision. This combination of machine learning expertise highlights the value of using diverse approaches to improve performance and reliability in addition to demonstrating the effectiveness of ensemble methods in handling challenging classification jobs.

There is much room for improvement and refinement in the integrated system. Further research into different models, feature engineering methods, and ensemble procedures may result in even higher performance improvements, boosting the system's effectiveness to unprecedented levels. In addition, continuous observation and assessment will be necessary to guarantee that the system continues to be flexible and sensitive to new risks and difficulties in the email environment. The integrated system is well-positioned to be at the forefront of spam detection technology by adopting culture of constant innovation and development. This will protect consumers from unsolicited email clutter and malicious threats in an increasingly digital environment.

ACKNOWLEDGMENT

We extend our heartfelt gratitude to Dr. S. C. Sharma, Director, AITR, Indore for his unwavering support and encouragement throughout the research process. His visionary leadership and invaluable guidance have been instrumental in shaping this research endeavor.

We would like to express our sincere appreciation to Dr. Kamal Kumar Sethi, Head of the Department of Computer Science and Engineering, for his insightful inputs and continuous encouragement.

We are also deeply grateful to Professor Krupi Saraf for her invaluable guidance, expertise, and mentorship throughout this research journey.

We acknowledge the collective efforts of all individuals who have contributed to this research project, directly or indirectly. Their support, encouragement, and collaboration have been indispensable in realizing our research goals.

Lastly, we extend our gratitude to the faculty, staff, and resources of the Department of Computer Science and Engineering whose unwavering support has been integral to the successful completion of this research work.

REFERENCES

[1] “A Bayesian Approach to Filtering Junk E-Mail” Mehran Sahami, Susan Dumaisy, David Heckermany and Eric Horvitzy.

[2] “Deep Learning for Email Spam Detection".Authors: Sustkova, K., & Matousek, M.

[3] S. K. Trivedi and S. Dey, "A Combining Classifiers Approach for Detecting Email Spams," *2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, Crans-Montana, Switzerland, 2016

[4] Q. Cheng, A. Xu, X. Li and L. Ding, "Adversarial Email Generation against Spam Detection Models through Feature Perturbation," *2022 IEEE International Conference on Assured Autonomy (ICAA)*, Fajardo, PR, USA, 2022

[5] “Improving Email Spam Filtering with Lightweight User Interaction and Reinforcement Learning" Authors: Carreras, X., & Marquez, L.

[6] A hybrid Data-Driven framework for Spam detection in Online Social Network Chanchal Kumara ,Taran Singh Bhartia , Shiv PrakashDepartment of Computer Science, Jamia Millia Islamia, New Delhi, India

[7] Email Spam Detection Using Naive Bayes Hrithik Vohra Dept. of Computer science & Engineering Delhi Technological University Delhi, India Manoj Kumar Dept. of Computer science & Engineering Delhi Technological University Delhi, India

[8] Content Based Spam Detection in Email using Bayesian Classifier Sunil B. Rathod, Tareek M. Pattewar

[9] https://www.kaggle.com/datasets/venky73/spam-mails-dataset

[10] https://researchonline.gcu.ac.uk/ws/portalfiles/portal/26991416/N.Mtetwa_Spam_Filtering_using_ML_ISCMI_Conference_Trimmed_Final.pdf

[11] https://www.ijste.org/articles/IJSTEV1I11008.pdf