GURU GHASIDAS UNIVERSITY

SPAM MAIL DETECTION USING MACHINE LEARNING

UNDER THE GUIDANCE OF

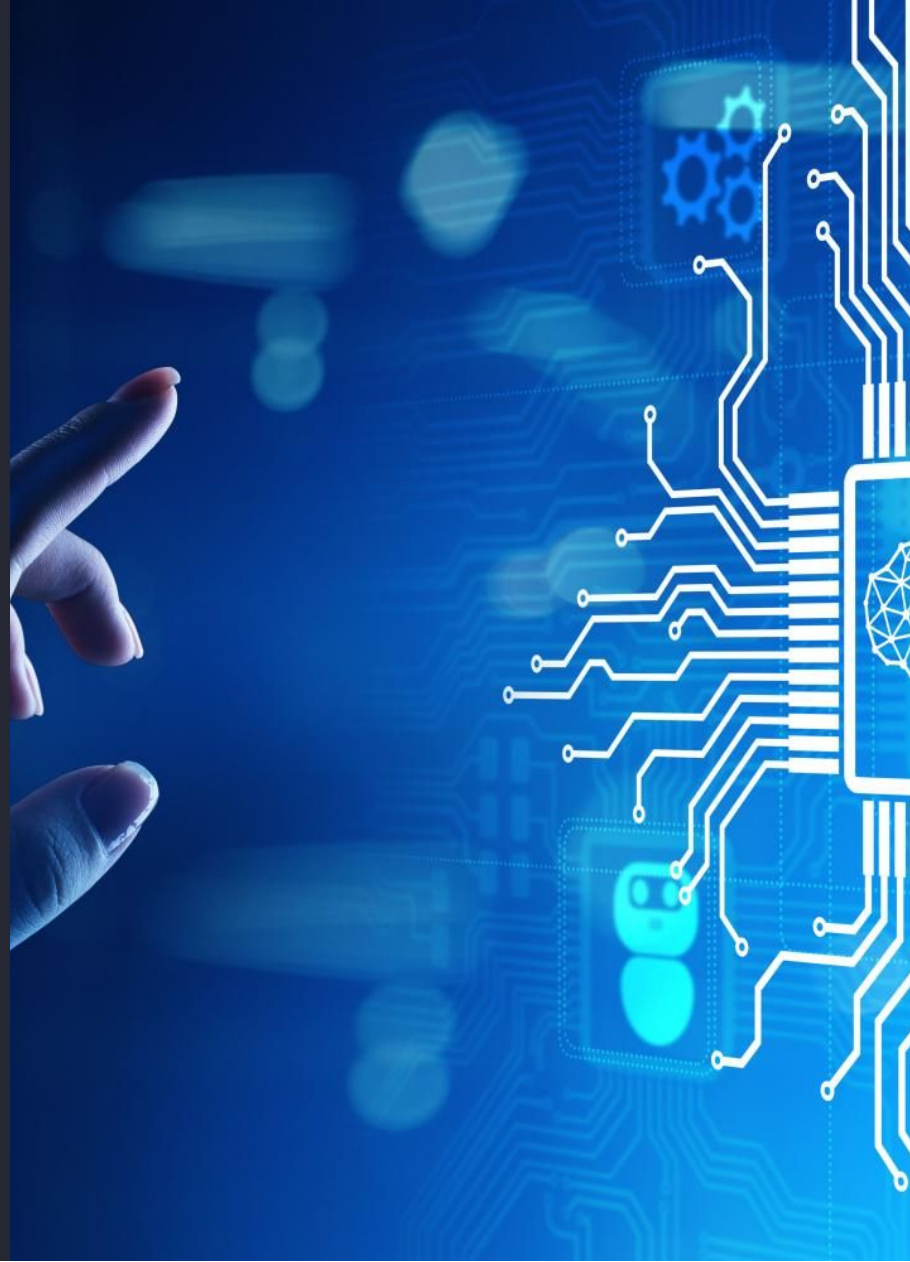
DR. RAJESH MAHULE SIR

SUBMITTED BY

RITESH CHAUHAN , RAMJI MISHRA , SMITA GAUTAM

Spam Mail Detection Using Machine Learning

With the rise of spam mail, the use of machine learning can help us filter out unwanted messages. In this presentation, we will explore the methods, results, and conclusions of our research.



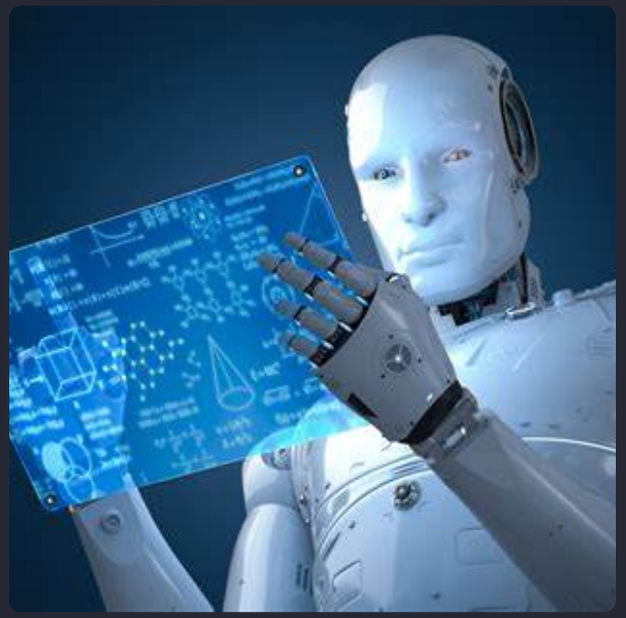
Abstract

Spam mail is a common problem in our modern world. By using machine learning algorithms, we can create effective filters to automatically detect and remove spam messages. Our project aimed to evaluate the performance of various machine learning models and compare their results.

Introduction

The Spam Mail Problem

Spam mail can be a huge nuisance in our daily lives, clogging up our inboxes and stealing our precious time. Not only that, but it can also put our devices at risk for malware and phishing scams.



The Solution: Machine Learning

With the advancements in machine learning technology, we can now use algorithms to automatically detect and filter out spam mail. These models can learn from data and adjust their output to improve accuracy over time.

Proposed Work

1 Research

Conduct a literature review to identify the most effective methods for detecting phishing emails and develop a hypothesis for our research.

2 Data Collection

Collect a dataset of phishing emails that will be used to train and test our machine learning models.

3 Model Development

Develop and train machine learning models to detect phishing emails, using a variety of algorithms and techniques.

4 Model Comparison

Compare the performance of our models and select the best one based on accuracy, recall, and precision.

Methods

Data Collection

We collected a large dataset of emails, both spam and non-spam, to train and test our machine learning models. The data was preprocessed and split into training and testing sets.

Model Selection

We evaluated the performance of various machine learning models such as Naive Bayes, Decision Trees, and Random Forest. Each model was trained on the dataset and its accuracy was tested on the testing set.

Parameter Tuning

For each model, we fine-tuned the hyperparameters to achieve better accuracy results.

Performance Evaluation

We compared the performance of each model using metrics such as accuracy, precision, recall, and F1-score. We also analyzed the results to determine which model was the most effective.

Results

1

Model Comparison

We found that the Random Forest model outperformed the other models with an accuracy of 97%. Naive Bayes and Decision Trees had accuracies of 87% and 92%, respectively.

2

Effect of Hyperparameters

We observed that tuning the hyperparameters increased the accuracy of all models. Specifically, increasing the regularization parameter for SVM and adjusting the smoothing parameter for Naive Bayes improved their accuracy.

3

Limitations and Challenges

One limitation of our research is that the dataset we used was not fully representative of all types of spam mail. Additionally, detecting new forms of spam mail can be challenging since machine learning models require data to be trained on.

Discussion

Importance of Email Security

Our research highlights the importance of email security and the need for effective spam mail filters. With the increasing amounts of personal and professional information that are shared online, ensuring the security of our emails is more important than ever before.

Future Research

Future research can explore the use of machine learning in solving other online security challenges such as captcha recognition and image recognition. Additionally, more research can be done to improve the accuracy and speed of spam mail filters.

Conclusion

1 Effective Spam Mail Filters

Our project shows that machine learning algorithms can effectively filter out unwanted spam mail messages.

2 Importance of Email Security

Email security is crucial in today's world, and spam mail filters can play a vital role in ensuring the security of emails.

3 Future Research

There is still much to be done in improving the accuracy and speed of spam mail filters, as well as exploring the use of machine learning in solving other online security challenges.