WWW.IJRPR.COM

International Journal of Research Publication and Reviews

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

DETECTION OF UNWANTED EMAIL MESSAGES

Dr. K. CHITRA[1], GOWTHAM .V[2]

[1]Research Scholar, Assistant Professor Department of Computer Science
[2]Department of Computer Science Sri Krishna Adithya College of Arts and Science Kovaipudur, Coimbatore, Tamil Nadu, India
gowthamkum7@gmail.com

ABSTRACT:

The detection of unwanted email messages, commonly known as spam, is a critical issue in cybersecurity. Spam emails are not only a nuisance but can also contain scams, malware, or phishing attempts. To protect users, organizations, and researchers are dedicated to developing robust filters for spam email detection.

Recent advancements in machine learning have led to the creation of highly effective spam filters. These filters are designed to analyze the content of emails and classify them as spam or legitimate based on various characteristics such as the sender's information, the presence of certain keywords, and patterns in the email's structure.

Despite the high performance of these filters, users continue to report an increase in frauds and attacks via spam emails. This indicates ongoing challenges in the field, such as the dynamic nature of spamming techniques and the presence of adversaries who continuously evolve their strategies to bypass filters.

Effective spam detection systems must be able to adapt to the ever-changing environment and consider the dataset shift problem, where the characteristics of spam emails change over time, leading to potential degradation in filter performance

Keywords: Tensor Flow-based spam detector might perform exploratory data analysis Exploratory Data Analysis (EDA)

Introduction :

In recent years, internet has become an integral part of life. With increased use of internet, numbers of email users are increasing day by day. This increasing use of email has created problems caused by Spam. In today's world, email is used in almost every industry, from business to education. Emails can be categorized into two categories: normal and spam. Junk emails, also known as spam messages, are emails that have been designed to harm recipients by wasting their time, computing resources, and stealing their valuable information. Spam detection is becoming a big challenge for network resources and users
because of some negative effects. Spam causes annoyance and wastes users time to regularly check and delete this large number of unwanted messages. It is estimated that spam emails are increasing at a rapid rate. One of the most important and prominent spam prevention techniques is filtering email. The aim of proposes work presented in this paper is to develop a spam detection algorithm that efficiently classifies a document into spam or non spam and to analyze how accurately they are classified into their original categories. A classifier is used to classify a mail to be either spam or non spam. Classification is a task of learning data patterns that are present in the data from the previous known instances and associating those data patterns with the classes.

OVERVIEW OF PROPOSED SYSTEM ALGORITHM :

the automatic filtering scheme has been applied in order to eliminate the unwanted messages in the social mail networks. The proposed system applies short text classifier (STC) in the extraction and selection of a set of characterizing and discriminate features.
The proposed system is a system used to detect unwanted messages in multitier web applications/mail . Our approach can create customized rule models of users. To achieve this, employ a lightweight virtualization technique to assign each user's filtering scheme to a dedicated container, an isolated virtual computing environment.

Exploratory Data Analysis (EDA) :

Exploratory Data Analysis (EDA) is a crucial step in the process of building a TensorFlow- based spam detector. EDA involves analyzing and visualizing data to uncover underlying patterns, spot anomalies, and test hypotheses.

Here's how EDA might be performed for a spam detector:

Data Collection: Gather a dataset of email or SMS messages labeled as spam or ham (non-spam).

Data Cleaning: Preprocess the data to remove any irrelevant information, correct errors, and handle missing values.

Visualization: Use plots and charts to visualize the distribution of spam and ham messages, the frequency of specific words, and other relevant characteristics.

Statistical Analysis: Calculate statistical measures like mean, median, mode, and standard deviation to understand the dataset's central tendencies and variability.

Feature Engineering: Identify and create new features that could be helpful for the spam detection model, such as the length of messages, the presence of certain keywords, or the use of capital letters.

Correlation Analysis: Determine if there are any correlations between different features and the likelihood of a message being spam.

Text Analysis: Perform text analysis to identify the most common words in spam and ham messages and create word clouds or frequency distributions.

Model Preparation: Prepare the data for modeling by splitting it into training and testing sets, encoding categorical variables, and normalizing or scaling numerical features.

Advantages

1. Preprocessing of Emails before actual storage.
2. Email classification can be applied to several different applications, including filtering messages based on priority, assigning messages to user-created folders, or identifying SPAM.
3. To ease the work and Reduce Time and Cost.
4. Mail Overflow: The amount of unwanted incoming mail can easily rise so much that it becomes an annoyance**.**

Conclusion :

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. Proposed system successfully implements The automatic filtering scheme has been applied in order to eliminate the unwanted messages in the social mail networks. The proposed system applies short text classifier (STC) in the extraction and selection of a set of characterizing and discriminate features.The proposed system is a system used to detect unwanted messages in multitier web applications/mail maintain in test data database.

REFERENCES BOOK :

1. Jon Meyer and Troy Downing, Java Virtual Machine, O'Reilly, 1997.
2. George Reese, Database Programming with JDBC and Java, O'Reilly, 1997.
3. Prashant Sridharan, Advanced Java Networking, Prentice-Hall, 1997.
4. John Zukowski, Java AWT Reference, O'Reilly, 1997.
5. Ken Arnold and James Gosling, The Java Programming Language, second ed., Addison- Wesley, 1998.

REFERENCES WEBSITE :

1. www.javatpoint.com
2. http://stackoverflow.com
3. www.codeproject.com
4. http://www.dzone.com
5. http://leetcode.com/