Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Email Spam: A Comprehensive Review of Optimize Detection Methods, Challenges, and Open Research Problems

EKRAMUL HAQUE TUSHER[1], MOHD ARFIAN ISMAIL[1,2], MD ARAFATUR RAHMAN[3],(Senior Member, IEEE),ALI H ALENEZI[4],MUEEN UDDIN[5],(Senior Member, IEEE)

[1]Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah, Pahang, 26600, Malaysia
[2]Center of Excellence for Artificial Intelligence & Data Science, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuhraya Tun Razak, Gambang, 26300, Malaysia
[3]School of Mathematics and Computer Science, University of Wolverhampton, WV1 1LY Wolverhampton, U.K.
[4]Remote Sensing Unit, Electrical Engineering Department. Northern Border University, Arar, Saudi Arabia.
[5]College of Computing and Information Technology, University of Doha for Science and Technology, Doha, Qatar

Corresponding author: Mohd Arfian Ismail (e-mail: arfian@umpsa.edu.my) and Mueen Uddin (mueen.uddin@udst.edu.qa)

This open access research is supported by Qatar National Library QNL and This study was supported by Fundamental Research Grant (FRGS) with FRGS/1/2022/ICT02/UMP/02/2 (RDU220134) from the Ministry of Higher Education Malaysia.

ABSTRACT Nowadays, emails are used across almost every field, spanning from business to education. Broadly, emails can be categorized as either ham or spam. Email spam, also known as junk emails or unwanted emails, can harm users by wasting time and computing resources, along with stealing valuable information. The volume of spam emails is rising rapidly day by day. Detecting and filtering spam presents significant and complex challenges for email systems. Traditional identification techniques like blacklists, real-time blackhole listing, and content-based methods have limitations. These limitations have led to the advancement of more sophisticated machine learning (ML) and deep learning (DL) methods for enhanced spam detection accuracy. In recent years, considerable attention has focused on the potential of ML and DL methods to improve email spam detection. A comprehensive literature review is therefore imperative for developing an updated, evidence-based understanding of contemporary research on employing these methods against this persistent problem. The review aims to systematically identify various ML and DL methods applied for spam detection, evaluate their effectiveness, and highlight promising future research directions considering gaps. By combining and analyzing findings across studies, it will obtain the strengths and weaknesses of existing methods. This review seeks to advance knowledge on reliable and efficient integration of state-of-the-art ML and DL into identifying email spam.

INDEX TERMS Email Spam, Machine Learning, Deep Learning, Fuzzy System, Feature Selection , Spam Detection.

I. INTRODUCTION

EMAILS have become an essential component of the contemporary lifestyle, which is heavily influenced by technology. Since its introduction to the public in the mid-1990s, the use of emails has had a noticeable positive effect on various sectors such as business, healthcare, education, and industry. Emails have facilitated collaboration among individuals by offering a cost-effective and expeditious mode of communication. [1]. They have greatly facilitated communication and information exchange on both personal and professional levels. However, the increasing usage and reliance on emails have also exposed users to greater cybersecurity risks in the form of spam attacks, malware infections, and other modes of exploitation [2]. As emails continue to play a pivotal role across domains, it is critical for users as well as organizations to adopt safe email practices and robust security measures against emerging threats. Cybercriminals utilize email channels as a launchpad for assaults that have the potential to seriously hurt both people