

# Spam Mail Detection using Machine Learning

Shubham Zanzad<sup>1</sup>, Devansh Thard<sup>2</sup>, Tushar Jarare<sup>3</sup>, Santosh Shinde<sup>4</sup>, Prof. B. S. Gayal

Department of Information Technology Engineering  
Sinhgad Academy of Engineering, Pune, Maharashtra, India

**Abstract:** *Email is now used in practically every industry, from business to education. Email is divided into subcategories, such as ham and spam. Unsolicited email, commonly known as spam or junk email, is a type of email that can be used to hurt consumers by wasting their time, wasting their computer resources, and stealing valuable data. Spam is increasing at an alarming rate every day. Spam detection and filtering have recently become major and widespread challenges for email and IoT service providers. Email filtration is among the most important and notable advanced approaches for detecting and preventing spam. Many machine learning and deep learning approaches, such as Nave Bayes, decision trees, neural networks, and random forests, have been utilised for this purpose. This article classifies utility research methodologies for based on machine learning tactics used in texting and IoT platforms into relevant classifications and analyses them. The accuracy, precision, memorability, and other characteristics of these techniques are all evaluated in depth. Finally, broad thoughts and research directions for the future are presented.*

**Keywords:** SVM, Decision Tree, K-Nearest Neighbor, Naïve Bayes

## REFERENCES

- [1]. Chao Chen, Yu Wang, Jun Zhang, Yang Xiang, Wanlei Zhou, Statistical Features-Based Real-Time Detection of Drifted Twitter Spam, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 12, NO. 4, APRIL 2017.
- [2]. L. Breiman, Random forests, Mach. Learn., vol. 45, no. 1, pp. 5-32, 2001.
- [3]. C. Grier, K. Thomas, V. Paxson, and M. Zhang, @spam: The underground on 140 characters or less, in Proc. 17th ACM Conf. Comput. Commun. Security, 2010, pp. 27-37.
- [4]. H. Kwak, C. Lee, H. Park, and S. Moon, What is twitter, a social network or a news media? in Proc. 19th Int. Conf. World Wide Web, 2010, pp. 591-600.
- [5]. K. Lee, J. Caverlee, and S. Webb, Uncovering social spammers: Social honeypots + machine learning, in Proc. 33rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., 2010, pp. 435-442.
- [6]. J. Oliver, P. Pajares, C. Ke, C. Chen, and Y. Xiang, An in-depth analysis of abuse on twitter, Trend Micro, Irving, TX, USA, Tech. Rep., Sep. 2014.
- [7]. Song, S. Lee, and J. Kim, Spam ltering in twitter using sender-receiver relationship, in Proc. 14th Int. Conf. Recent Adv. Intrusion Detection, 2011, pp. 301-317.
- [8]. K. Thomas, C. Grier, D. Song, and V. Paxson, Suspended accounts in retrospect: An analysis of twitter spam, in Proc. ACM SIGCOMM Conf. Internet Meas. Cof., 2011, pp. 243-258.
- [9]. C. Yang, R. Harkreader, and G. Gu, Empirical evaluation and new design for fighting evolving twitter spammers, IEEE Trans. Inf. Forensics Security, vol. 8, no. 8, pp. 1280-1293, Aug. 2013.
- [10]. S. Yardi, D. Romero, G. Schoenebeck, and D. Boyd, Detecting spam in a twitter network, First Monday, vol. 15, nos. 1-4, pp. 1-13, Jan. 2010.