

Spam Detection and Fake User Identification in Twitter: An Analysis of Machine Learning Models

Radhika S¹, Phani Sri Siva Charan K², Sai Khush Kumar M³,
Ramanjaneya Reddy K⁴, Manish Prasad⁵

Associate Professor, Department of Computer Science and Engineering¹
Students, Department of Computer Science and Engineering^{2,3,4,5}
Raghu Institute of Technology, Visakhapatnam, AP, India

Abstract: *Spammer detection and fake user identification are significant issues in the realm of social media, with Twitter being no exception. The detection of spammers and fake users on Twitter is essential for preserving the platform's integrity and protecting users from online scams and fraud. This project paper aims to conduct a comprehensive study of the different techniques and algorithms used for spammer detection and fake user identification on Twitter. We will evaluate the effectiveness of traditional techniques and machine learning-based methods and propose a novel approach combining both methods for spammer detection and fake user identification on Twitterpils.*

Keywords: Social Networks, Spammer Detection, Fake User Identification, Machine Learning, Decision Trees, Random Forests.

REFERENCES

- [1]. B. Erçahin, Ö. Aktaş, D. Kiliç, and C. Akyol, "Twitter fake account detection," in Proc. Int. Conf. Comput. Sci. Eng. (UBMK), Oct. 2017, pp. 388–392.
- [2]. F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, "Detecting spammers on Twitter," in Proc. Collaboration, Electron. Messaging, Anti- Abuse Spam Conf. (CEAS), vol. 6, Jul. 2010, p. 12.
- [3]. S. Gharge, and M. Chavan, "An integrated approach for malicious tweets detection using NLP," in Proc. Int. Conf. Inventive Commun. Comput. Technol. (ICICCT), Mar. 2017, pp. 435–438.
- [4]. T. Wu, S. Wen, Y. Xiang, and W. Zhou, "Twitter spam detection: Survey of new approaches and comparative study," Comput. Secur., vol. 76, pp. 265–284, Jul. 2018.
- [5]. S. J. Soman, "A survey on behaviors exhibited by spammers in popular social media networks," in Proc. Int. Conf. Circuit, Power Comput. Tech- nol. (ICCPCT), Mar. 2016, pp. 1–6.
- [6]. A. Gupta, H. Lamba, and P. Kumaraguru, "1.00 per RT #BostonMarathon # prayforboston: Analyzing fake content on Twitter," in Proc. eCrime Researchers Summit (eCRS), 2013, pp. 1–12.
- [7]. F. Concone, A. De Paola, G. Lo Re, and M. Morana, "Twitter analysis for real-time malware discovery," in Proc. AEIT Int. Annu. Conf., Sep. 2017, pp. 1–6.
- [8]. N. Eshraqi, M. Jalali, and M. H. Moattar, "Detecting spam tweets in Twitter using a data stream clustering algorithm," in Proc. Int. Congr. Technol., Commun. Knowl. (ICTCK), Nov. 2015, pp. 347–351.
- [9]. C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou, and G. Min, "Statistical features-based real-time detection of drifted Twitter spam," IEEE Trans. Inf. Forensics Security, vol. 12, no. 4, pp. 914–925, Apr. 2017.
- [10]. C. Buntain and J. Golbeck, "Automatically identifying fake news in popular Twitter threads," in Proc. IEEE Int. Conf. Smart Cloud (SmartCloud), Nov. 2017, pp. 208–215.
- [11]. C. Chen, J. Zhang, Y. Xie, Y. Xiang, W. Zhou, M. M. Hassan, A. AlElaiwi, and M. Alrubaian, "A performance evaluation of machine learning-based streaming spam tweets detection," IEEE Trans. Comput. Social Syst., vol. 2, no. 3, pp. 65–76, Sep. 2015.
- [12]. G. Stafford and L. L. Yu, "An evaluation of the effect of spam on Twitter trending topics," in Proc. Int. Conf. Social Comput., Sep. 2013, pp. 373–378.

- [13]. M. Mateen, M. A. Iqbal, M. Aleem, and M. A. Islam, "A hybrid approach for spam detection for Twitter," in Proc. 14th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST), Jan. 2017, pp. 466–471.
- [14]. A. Gupta and R. Kaushal, "Improving spam detection in online social networks," in Proc. Int. Conf. Cogn. Comput. Inf. Process. (CCIP), Mar. 2015, pp. 1–6.
- [15]. F. Fathaliani and M. Bouguessa, "A model-based approach for identifying spammers in social networks," in Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA), Oct. 2015, pp. 1–9.
- [16]. V. Chauhan, A. Pilaniya, V. Middha, A. Gupta, U. Bana, B. R. Prasad, and S. Agarwal, "Anomalous behavior detection in social networking," in Proc. 8th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT), Jul. 2017, pp. 1–5.
- [17]. S. Jeong, G. Noh, H. Oh, and C.-K. Kim, "Follow spam detection based on cascaded social information," Inf. Sci., vol. 369, pp. 481–499, Nov. 2016.
- [18]. M. Washha, A. Qaroush, and F. Sedes, "Leveraging time for spammers detection on Twitter," in Proc. 8th Int. Conf. Manage. Digit. EcoSyst., Nov. 2016, pp. 109–116.
- [19]. B. Wang, A. Zubiaga, M. Liakata, and R. Procter, "Making the most of tweet-inherent features for social spam detection on Twitter," 2015, arXiv:1503.07405. [Online]. Available: <https://arxiv.org/abs/1503.07405>
- [20]. M. Hussain, M. Ahmed, H. A. Khattak, M. Imran, A. Khan, S. Din,
- [21]. A. Ahmad, G. Jeon, and A. G. Reddy, "Towards ontology-based multilingual URL filtering: A big data problem," J. Supercomput., vol. 74, no. 10, pp. 5003–5021, Oct. 2018.
- [22]. B. Meda, E. Ragusa, C. Gianoglio, R. Zunino, A. Ottaviano, E. Scillia, and
- [23]. R. Surlinelli, "Spam detection of Twitter traffic: A framework based on random forests and non-uniform feature sampling," in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), Aug. 2016, pp. 811–817.
- [24]. S. Ghosh, G. Korlam, and N. Ganguly, "Spammers' networks within online social networks: A case-study on Twitter," in Proc. 20th Int. Conf. Companion World Wide Web, Mar. 2011, pp. 41–42.
- [25]. C. Chen, S. Wen, J. Zhang, Y. Xiang, J. Oliver, A. Alelaiwi, and M. M. Hassan, "Investigating the deceptive information in Twitter spam," Future Gener. Comput. Syst., vol. 72, pp. 319–326, Jul. 2017.
- [26]. David, O. S. Siordia, and D. Moctezuma, "Features combination for the detection of malicious Twitter accounts," in Proc. IEEE Int. Autumn Meeting Power, Electron. Comput. (ROPEC), Nov. 2016, pp. 1–6.
- [27]. M. Babcock, R. A. V. Cox, and S. Kumar, "Diffusion of pro- and anti-false information tweets: The black panther movie case," Comput. Math. Org. Theory, vol. 25, no. 1, pp. 72–84, Mar. 2020.
- [28]. S. Keretna, A. Hossny, and D. Creighton, "Recognising user identity in Twitter social networks via text mining," in Proc. IEEE Int. Conf. Syst., Man, Cybern., Oct. 2013, pp. 3079–3082.
- [29]. C. Meda, F. Bisio, P. Gastaldo, and R. Zunino, "A machine learning approach for Twitter spammers detection," in Proc. Int. Carnahan Conf. Secur. Technol. (ICCST), Oct. 2014, pp. 1–6.
- [30]. W. Chen, C. K. Yeo, C. T. Lau, and B. S. Lee, "Real-time Twitter content polluter detection based on direct features," in Proc. 2nd Int. Conf. Inf. Sci. Secur. (ICISS), Dec. 2015, pp. 1–4.
- [31]. H. Shen and X. Liu, "Detecting spammers on Twitter based on content and social interaction," in Proc. Int. Conf. Netw. Inf. Syst. Comput., pp. 413–417, Jan. 2015.
- [32]. G. Jain, M. Sharma, and B. Agarwal, "Spam detection in social media using convolutional and long short term memory neural network," Ann. Math. Artif. Intell., vol. 85, no. 1, pp. 21–44, Jan. 2020.
- [33]. [31] M. Washha, A. Qaroush, M. Mezghani, and F. Sedes, "A topic-based hidden Markov model for real-time spam tweets filtering," Procedia Comput. Sci., vol. 112, pp. 833–843, Jan. 2017.
- [34]. F. Pierrri and S. Ceri, "False news on social media: A data-driven survey," 2020, arXiv:1902.07539. [Online]. Available: <https://arxiv.org/abs/1902.07539>
- [35]. S. Sadiq, Y. Yan, A. Taylor, M.-L. Shyu, S.-C. Chen, and D. Feaster, "AAFA: Associative affinity factor analysis for bot detection and stance classification in Twitter," in Proc. IEEE Int. Conf. Inf. Reuse Integr. (IRI), Aug. 2017, pp. 356–365.

- [36]. M. U. S. Khan, M. Ali, A. Abbas, S. U. Khan, and A. Y. Zomaya, “Segregating spammers and unsolicited bloggers from genuine experts on Twitter,” IEEE Trans. Dependable Secure Comput., vol. 15, no. 4, pp. 551–560, Jul./Aug. 2018.