Preventing Poisoning Attacks On AI Based Threat Intelligence Systems

Submitted by grigby1 on Wed, 11/04/2020 - 2:13pm

Title	Preventing Poisoning Attacks On AI Based Threat Intelligence Systems
Publication Type	Conference Paper
Year of Publication	2019
Authors	Khurana, N., Mittal, S., Piplai, A., Joshi, A.
Conference Name	2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)
Date Published	Oct. 2019
Publisher	IEEE
ISBN Number	978-1-7281-0824-7
Keywords	AI Poisoning, AI systems, artificial intelligence, computer security, cybersecurity domain, Engines, ensembled semi-supervised approach, Human Behavior, learning (artificial intelligence), malicious information, online social media, poisoning attacks prevention, pubcrawl, resilience, Resiliency, Scalability, security analysts, security of data, social networking (online), Support vector machines, threat intelligence systems, Twitter, Web sites

Abstract

As AI systems become more ubiquitous, securing them becomes an emerging challenge. Over the years, with the surge in online social media use and the data available for analysis, AI systems have been built to extract, represent and use this information. The credibility of this information extracted from open sources, however, can often be questionable. Malicious or incorrect information can cause a loss of money, reputation, and resources; and in certain situations, pose a threat to human life. In this paper, we use an ensembled semi-supervised approach to determine the credibility of Reddit posts by estimating their reputation score to ensure the validity of information ingested by AI systems. We demonstrate our approach in the cybersecurity domain, where security analysts utilize these systems to determine possible threats by analyzing the data scattered on social media websites, forums, blogs, etc.

URL	https://ieeexplore.ieee.org/document/8918803
DOI	10.1109/MLSP.2019.8918803

Citation Key khurana_preventing_2019

🏷 AI Poisoning AI systems Artificial Intelligence computer security cybersecurity domain Engines ensembled semi-supervised approach Human behavior learning (artificial intelligence) malicious information online social media poisoning attacks prevention pubcrawl resilience Resiliency Scalability security analysts security of data social networking (online) Support vector machines threat intelligence systems Twitter Web sites