MACHINE LEARNING AND IT'S SECURITY RISKS

M Mahbubul Alam[*], *Md.Sarowar Alam Sagor*[*], *Saifur Rahman*[*]

[*]Dept. of Computer Science & Engineering
Rajshahi University of Engineering & Technology, Rajshahi, Bangladesh

ABSTRACT

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves. In dealing with evolving input, Machine learning systems offer unparallel flexibility such as intrusion detection systems and spam email filtering. But machine learning algorithms themselves can be target of attack by a malicious adversary. This paper includes a categorization of different types of attacks on machine learning systems and technique. Three axes may define a space of attacks. These are influence, specificity and security violation. The Paper also provides a variety of defenses against those attacks such as increasing robustness, detecting attacks, confusing the adversary with disinformation, randomization for the targeted attacks. In the end, it will provide a discussion of ideas that are important to security for machine learning , a model which will give a lower bound on attacker's work function and a list of open problems. Machine learning will be used in robotics, auto mobile, biomatrix and in all of the field in near future which will change the face of the world. It's very important to ensure it's security.