SIT719 Security and Privacy Issues in Analytics

Pass Task 7.1: Taxonomy of Attacks, Defenses, and Consequences in Adversarial Machine Learning

Overview

The Information Technology Laboratory (ITL) at the National Institute of Standards and Technology (NIST) promotes the U.S. economy and public welfare by providing technical leadership for the Nation's measurement and standards infrastructure. Recently NIST has published an internal report on "A Taxonomy and Terminology of Adversarial Machine Learning" (link below). This NIST Interagency/Internal Report (NISTIR) is intended as a step toward securing applications of Artificial Intelligence (AI), especially against adversarial manipulations of Machine Learning (ML), by developing a taxonomy and terminology of Adversarial Machine Learning (AML).

Link: https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8269-draft.pdf

Please see the details of the task in the Task Description section.

This is a *Pass task*, so you MUST complete the task and submit the evidence of your work to Ontrack.

Task Description

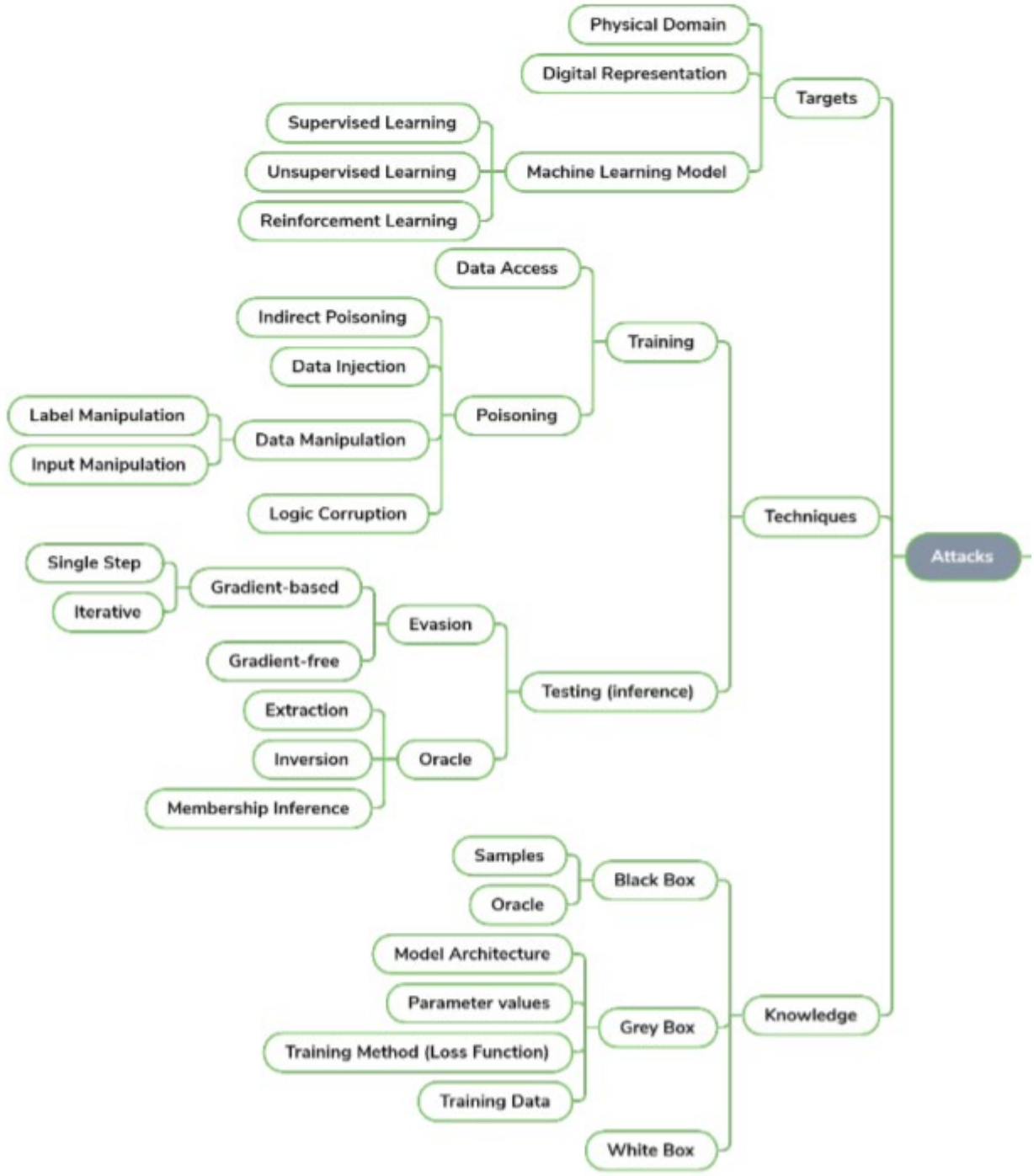
Suppose you are working in an organization who are developing a report on the vulnerabilities of machine learning models due to adversarial attacks. Your manager has asked you to provide a 600 word report to submit within the next week. His expectation is that the 600 word report will cover the attack taxonomies, defense mechanisms and consequences.

Instructions:

1. Read the NIST article from the below link:

https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8269-draft.pdf

2. Identify five important attack types. Summarize in approx. 300 words.



Hint: The above figure demonstrates the attack categories. It has been obtained from Figure 2 of the report.

2. Summarize the defense mechanisms for the attack types you identified in step 1. (Approx. 200 words)

Submit the report PDF to the OnTrack system.