Explainable Artificial Intelligence in CyberSecurity:

ABSTRACT

Nowadays, Artificial Intelligence (AI) is widely applied in every area of a human being's daily life. Despite the AI benefits, its application suffers from the opacity of complex internal mechanisms and doesn't satisfy by design the principles of Explainable Artificial Intelligence (XAI). The lack of transparency further exacerbates the problem in the field of CyberSecurity because entrusting crucial decisions to a system that cannot explain itself presents obvious dangers. There are several methods in the literature capable of providing explainability of AI results. Anyway, the application of XAI in CyberSecurity can be a double-edged sword. It substantially improves the CyberSecurity practices but simultaneously leaves the system vulnerable to adversary attacks. Therefore, there is a need to analyze the state-of-the-art of XAI methods in CyberSecurity to provide a clear vision for future research. This study presents an in-depth examination of the application of XAI in CyberSecurity. It considers more than 300 papers to comprehensively analyze the main CyberSecurity application fields, like Intrusion Detection Systems, Malware detection, Phishing and Spam detection, BotNets detection, Fraud detection, Zero-Day vulnerabilities, Digital Forensics and Crypto-Jacking. Specifically, this study focuses on the explainability methods adopted or proposed in these fields, pointing out promising works and new challenges.

Guide: *Sreedeepa*

Name: Mohammad Ali

Roll No: *33*