RADemics

Explainable Artificial Intelligence Techniques for Enhancing Trust in Predictive Security Models

B.Lakshmi Dhevi, T R Vedhavathy , Mohammed Muzaffar Hussain

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY, SRM INSTITUTE OF SCIENCE AND TECHNOLOGY, C ABDUL HAKEEM COLLEGE OF ENGINEERING AND TECHNOLOGY.

.

11 Explainable Artificial Intelligence Techniques for Enhancing Trust in Predictive Security Models

1B.Lakshmi Dhevi, Department of Networking and Communication ,SRM Institute of Science and Technology, College of Engineering and Technology, Kattankulathur, Chennai. lakshmidhevi.b@gmail.com

2T R Vedhavathy, AP/NWC, SRM Institute of Science and Technology, Kattankulathur. trveda@gmail.com

3Mohammed Muzaffar Hussain, Professor , Department of AI&DS
C Abdul Hakeem College of Engineering and Technology, mhd.muzaffar@gmail.com

Abstract

Explainable Artificial Intelligence (XAI) has emerged as a crucial component in enhancing the transparency, trust, and effectiveness of predictive security models. As cybersecurity threats become increasingly sophisticated, the need for interpretable models that ensure reliable decision-making in real-time has never been greater. This chapter explores the integration of XAI techniques in security systems, focusing on their role in improving model interpretability without compromising predictive accuracy. The challenges of balancing computational efficiency with high-quality explanations in real-time anomaly detection systems are discussed, alongside the importance of feature importance analysis for understanding model behavior. Model-agnostic approaches such as SHAP and LIME are examined for their ability to provide insights into complex, black-box models while maintaining high levels of accuracy. The chapter further investigates the evolving landscape of cybersecurity, where dynamic and novel threats require continuous adaptation of anomaly detection systems. It highlights the practical implications of XAI in fostering trust and providing actionable insights for security professionals, ensuring that security measures are both robust and understandable. The chapter concludes by identifying future directions for XAI research in predictive security models, emphasizing the need for scalable, efficient, and secure explainability solutions.

Keywords: Explainable Artificial Intelligence, Predictive Security Models, Anomaly Detection, Model Interpretability, Feature Importance, Cybersecurity.

Introduction

Explainable Artificial Intelligence (XAI) has become an indispensable aspect of modern machine learning applications, particularly in cybersecurity, where trust and transparency are paramount [1]. Predictive security models powered by AI have revolutionized the way threats are detected, but their "black-box" nature often raises concerns regarding the rationale behind their predictions [2]. In critical fields such as cybersecurity, where decisions can significantly impact organizational security, understanding how a model arrives at a decision was crucial [3]. The lack of transparency in AI systems often undermines the confidence of users and stakeholders, making

it essential to introduce explainability techniques that bridge this gap [4]. This chapter explores the intersection of XAI and predictive security models, discussing the role of explainable AI in enhancing the trust and reliability of security systems while preserving predictive accuracy [5].

As cybersecurity systems evolve to address increasingly sophisticated threats, the demand for models that can detect anomalies in real-time has surged [6]. Traditional machine learning models, although highly effective, often operate as opaque "black boxes" that do not provide explanations for their outputs [7]. This lack of interpretability poses challenges in high-stakes environments where professionals need to understand the rationale behind security decisions [8]. XAI techniques, such as LIME (Local Interpretable Model-agnostic Explanations) and SHAP (Shapley Additive Explanations), have been developed to address this issue [9]. These methods allow users to gain insights into the decision-making process of AI systems, providing transparency without compromising the model's performance [10]. By making the inner workings of predictive models more understandable, XAI fosters greater trust in AI-driven security systems, which was essential for their widespread adoption [11].

The integration of XAI into predictive security models presents unique challenges, particularly in real-time anomaly detection systems [12]. The primary obstacle lies in balancing the need for interpretability with the requirement for quick decision-making [13]. In cybersecurity, where timely responses are critical, the computational cost of generating explanations can introduce delays [14]. For instance, real-time anomaly detection models must be able to process vast amounts of data and identify potential threats instantaneously [15]. Traditional XAI techniques often require additional computational resources to generate explanations, which can slow down the overall performance of the system [16]. As a result, finding ways to apply XAI methods without hindering the real-time capabilities of predictive security models was one of the central challenges addressed in this chapter [17].