RADemics

Ensemble Learning Models for Improved Threat Prediction Accuracy

V. Shoba, Syed Naimatullah Hussain, N Legapriyadharshini

SRM Arts And Science College, Nagarjuna college of engineering and technology, ,Saveetha College of Liberal Arts and Sciences

Ensemble Learning Models for Improved Threat Prediction Accuracy

[1]V. Shoba, Assistant Professor, Computer Applications And Technology, SRM Arts And Science College, shobacs@srmasasc.ac.in

[2]Syed Naimatullah Hussain, professor & HoD, CSE (Data Science), Nagarjuna college of engineering and technology, near Devanahalli Bengaluru, Mail ID: syed.hussain@gmail.com

[3]N Legapriyadharshini, Associate Professor, Computer Science, Saveetha College of Liberal Arts and Sciences, SIMATS Chennai, legapriya.scals@saveetha.com

Abstract

The rapid evolution of cyber threats has necessitated the development of advanced machine learning techniques to ensure robust, scalable, and accurate threat detection in real-time. This book chapter explores the integration of hybrid models, specifically the fusion of ensemble learning methods and deep learning techniques, as a transformative approach for enhancing cybersecurity defenses. Hybrid models combine the strengths of diverse machine learning paradigms to address challenges such as high-dimensional data, imbalanced datasets, and the detection of previously unseen attack patterns. The chapter examines various ensemble methods, including bagging, boosting, and stacking, alongside deep learning architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). Real-world case studies highlight the effectiveness of these hybrid systems in diverse cybersecurity applications, including malware detection, network intrusion detection, and phishing prevention. The synergistic combination of deep learning's feature extraction capabilities with ensemble techniques' robustness significantly improves prediction accuracy, generalization, and scalability. Moreover, the ability of hybrid models to adapt to emerging and evolving threats positions them as a key component in the future of cybersecurity. This chapter provides comprehensive insights into the design, application, and performance evaluation of hybrid models, offering valuable knowledge for researchers, engineers, and practitioners seeking to strengthen cybersecurity infrastructures.

Keywords: Hybrid Models, Ensemble Learning, Deep Learning, Malware Detection, Network Intrusion Detection, Phishing Prevention.

Introduction

The rapid advancement of digital technologies has brought about significant improvements in global connectivity and productivity, yet it has simultaneously created new and more complex avenues for cyber threats to emerge [1]. As businesses and individuals increasingly rely on interconnected systems, the frequency and sophistication of cyber-attacks have escalated dramatically [2]. Traditional security systems, based on signature-based methods or simple rule-based detection, are increasingly inadequate for defending against these advanced, evolving threats [3]. This inadequacy has driven the need for more adaptive and intelligent solutions capable of detecting novel attack patterns in real-time [4]. In this context, machine learning (ML) and, more specifically, hybrid models that combine ensemble learning with deep learning techniques have gained prominence as an effective approach to enhancing cybersecurity defenses [5].

Ensemble learning methods, such as bagging, boosting, and stacking, have long been recognized for their ability to improve prediction accuracy and generalize well across different datasets [6]. These techniques involve combining multiple individual models to create a stronger, more reliable overall system [7]. In cybersecurity, ensemble models help address issues such as overfitting, high variance, and the challenge of detecting rare or previously unseen cyber-attacks [8]. However, while ensemble models provide robustness, they often lack the capacity to automatically extract meaningful features from complex, high-dimensional cybersecurity data [9]. This is where deep learning techniques, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), come into play. These models excel in learning hierarchical features from raw data and can capture intricate patterns in time-series data, such as network traffic or malware behavior [10].

The combination of ensemble learning and deep learning into hybrid models addresses the limitations of both approaches [11]. Ensemble techniques enhance the stability and accuracy of deep learning models, while deep learning provides ensemble methods with powerful feature extraction capabilities [12]. Hybrid models are therefore particularly suited to tackle the challenges inherent in cybersecurity, such as data imbalance, high-dimensional feature spaces, and real-time threat detection [13]. By aggregating the predictions of multiple deep models trained on diverse data subsets, hybrid systems can improve classification accuracy, mitigate the risk of overfitting, and ensure better generalization across various types of cyber threats [14]. This makes hybrid models not only more robust but also more adaptive to the constantly evolving nature of cyber-attacks [15].