A Comparative Analysis of Machine Learning Algorithms in Predictive Analytics for Cybersecurity Breach Detection

Cameron Jackson
PhD
Technical University of Munich
Arcisstraße 21, 80333 Munich, Germany

Kim Lee
Associate Professor
University of Waterloo
200 University Ave W, Waterloo, ON N2L 3G1, Canada

Kim Mitchell
Professor
University of Melbourne
Parkville, VIC 3010, Australia

Abstract. In the ever-evolving landscape of cybersecurity, the prevalence of data breaches necessitates robust predictive analytics to preemptively identify potential threats. This study conducts a comparative analysis of various machine learning algorithms, including Support Vector Machines, Random Forest, and Neural Networks, assessing their efficacy in breach detection. Through a systematic evaluation utilizing real-world datasets, we quantify the algorithms' performance based on accuracy, precision, and recall metrics. Our findings reveal that while Neural Networks offer superior accuracy, Random Forest excels in interpretability and speed. This research underscores the significance of selecting the appropriate algorithm in enhancing cybersecurity measures, providing valuable insights for practitioners and researchers alike.

Keywords: Machine Learning, Predictive Analytics, Cybersecurity, Data Breach Detection, Performance Evaluation, Support Vector Machines, Random Forest, Neural Networks

Introduction

Cybersecurity remains one of the most pressing challenges for organizations worldwide, with data breaches increasingly becoming a common occurrence. As businesses and governmental bodies transition further into the digital realm, the complexity and scale of cyber threats grow, demanding advanced and proactive approaches to threat detection and response. The cost of data breaches is staggering, affecting millions of users and resulting in significant financial and reputational damage. According to recent studies, the global cost of cybercrime is expected to reach $10.5 trillion annually by 2025, highlighting the urgent need for effective methodologies in breach detection. Machine learning (ML) has emerged as a pivotal technology in the field of cybersecurity, offering innovative solutions

that enhance the capability of existing security frameworks. This paper sets out to analyze the comparative efficiency of various ML algorithms in predictive analytics tailored for cybersecurity breach detection. We aim to evaluate the performance of well-established machine learning techniques, such as Support Vector Machines (SVM), Random Forest (RF), and Neural Networks (NN). By deploying these algorithms on diverse datasets encompassing varied breach scenarios, we establish a robust framework for assessing their effectiveness based on critical parameters such as accuracy, precision, and computational efficiency. Furthermore, the research also explores the interpretability of these algorithms, a crucial aspect when deploying machine learning solutions in real-world cybersecurity incidents. Clear identification and understanding of decision-making processes in breach detection models can greatly influence trust and acceptance among stakeholders. The paper is structured as follows: Section 2 provides a comprehensive overview of the existing literature on machine learning applications in cybersecurity. Section 3 details the datasets and methodologies employed in our experiments. Section 4 presents the comparative analysis of the algorithms, while Section 5 discusses the implications of our findings and offers recommendations for future research directions.

References

1. Рагимов, Э. Р. О. (2012). Методология оптимальной идентификации расположения программных единиц в комплексе безопасных программ, реализующих систему защиты информации корпоративной сети. Вопросы защиты информации, (1), 51-57.
2. Рагимов, Э. Р. (2011). Роль программных комплексов в управлении безопасностью корпоративной информационной системы. Телекоммуникации, (12), 4-7.
3. Rasif, R. E. (2010). BASE PRINCIPAL OF MANAGING OF NETWORK SOFTWARE SECURITY BY VULNERABILITIES DETERMINATION MODEL. Computer Science & Telecommunications, 28(5).
4. Рагимов, Э. Р. (2010). Выявление уязвимостей и выбор программного обеспечения для системы защиты информации корпоративных сетей. Телекоммуникации, (1), 46-48.
5. Алгулиев, Р. М., & Рагимов, Э. Р. (2006). Синтез архитектуры защищенной корпоративной сети. Телекоммуникации, (2), 19-23.
6. Рагимов, Э. Р. (2005). КЛАССИФИКАЦИЯ УГРОЗ ПО СЕРВИСАМ БЕЗОПАСНОСТИ НА ОСНОВЕ СТАТИСТИЧЕСКИХ ДАННЫХ. Информационное противодействие угрозам терроризма, (5), 87-92.
7. Рагимов, Э. Р. (2011). Модель идентификации безотказной работы программных средств в корпоративных сетях. Телекоммуникации, (2), 2-5.
8. Алгулиев, Р. М., & Рагимов, Э. Р. (2005). Об одном методе оценки информационной безопасности корпоративных сетей в стадии их проектирования. Информационные технологии, (7), 35-39.

9. Rahimov, E. (2007). TECHNICAL ASPECTS OF CENTRALIZING ADMINISTRATING OF MODERN CORPORATE NETWORKS SERVICES. ITTC–2007, 68.
10. Рагимов, Э. Р. (2009). Роль безопасности программного обеспечения в комплексной системе защиты корпоративных сетей. Телекоммуникации, (10), 23-26.
11. D. Satyanarayana, Sathyashree “A Three Level Architecture for Wireless Communication using LiFi”, in Recent Advances in Information and Communication Technology, Vol: 566, pp: 212-221, Springer International Publishing, 2018 book. Print ISBN: 978-3-319-60662-0; Electronic ISBN: 978-3-319-60663-7; Book Series.
12. Satyanarayana, D. and Rao, S. V. (2011), “Fault-tolerant spanners for ad hoc networks” International Journal of Network Management, Vol: 22, Issue: 4, Pages: 311-329, July-2012, Wiley Publishers. https://doi.org/10.1002/nem.807
13. D. Satyanarayana, Sathyashree. S “Mobility Management in Voronoi based Cell Structure for Wireless networks”, IEEE proceedings for the International Conferences on Currents trends in Information Technology (CTIT), Pages: 204-208, Dubai, December-2013. DOI: 10.1109/CTIT.2013.6749504; INSPEC Accession Number: 14131032