Journal of Artificial Intelligence and Big Data Disciplines (JAIBDD)

International | Peer Reviewed | Open Access | Online

Machine Learning Algorithms for Optimizing Big Data-Enhanced Cybersecurity in ERP Ecosystems

Shakir Syed,

AI & Analytics Leader,
Purdue University,
Bargersville, IN, USA

Abstract : The advent of big data has had a significant impact on enterprise resource planning (ERP) ecosystems, particularly when it comes to supporting scalability and addressing the limitations of existing cybersecurity frameworks in ERP ecosystems. Big data technologies enhance cybersecurity in ERP ecosystems by improving cyber forensics readiness. However, the use of big data-enhanced cybersecurity solutions in ERP ecosystems can result in several cybersecurity concerns regarding the privacy, protection, and preservation of ERP cybersecurity data. This creates a need for machine learning-based algorithms to optimize data analytics-based cybersecurity initiatives in ERP ecosystems. Five machine learning algorithms are developed to optimize ERP big data-enhanced cybersecurity.

In developing these machine learning models, a framework has been created for securing ERP big data by identifying and selecting the most appropriate machine learning algorithm that can be utilized to develop an effective ERP cybersecurity solution. Each of the algorithms has been analyzed in terms of its evaluation metrics and other performance and learning attributes. While all the algorithms can be effectively used for ERP big data-enhanced cybersecurity, the following are the outstanding strengths of each algorithm: logistic regression for ensuring scalability and making classification predictions based on real-time assessments; decision tree for easily integrating with existing ERP systems; random forest for its ensemble learning-based power to enhance overall ERP ecosystem security; k-nearest neighbors for its simple and easy-to-understand methodology; and support vector machine for its potential in ERP system security data clustering and addressing the system security multi-dimensionality challenge. This paper also presents the limitations of each evaluated algorithm, as well as areas for additional research to maximize the overall effectiveness of the machine learning techniques.

Key words : Big Data, ERP Ecosystems, Scalability, Cybersecurity Frameworks, Cyber Forensics, Data Privacy, Machine Learning, Data Protection, Cybersecurity Optimization, Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors, Support Vector Machine, Real-Time Assessments, Ensemble Learning, Data Clustering, Multi-Dimensionality, Algorithm Evaluation, Research Areas.

1. Introduction

Enterprise Resource Planning (ERP) systems manage corporate information and are used to optimize and track a multitude of internal business processes. ERP systems continue to enforce strict security measures since their data is mission-critical. Big data analytics tools are a strong cybersecurity complement as they enable deterministic, real-time, proactive, and predictive security measures. The big data preventive approach enhances real-time security through the establishment of likelihood thresholds and the active intervention of governance, risk, and compliance personnel that set corporate policies on cross-application privileges and monitor user behaviors compared to usual baseline behaviors to prevent any fraudulent data extraction from taking place. The efficacy of machine learning-based algorithms integrated with the existing protective tools and business intelligence security enforcement component products of an advanced ERP system to optimize cybersecurity intelligence is reported. The machine learning-driven Bayesian algorithm and its supporting machine learning nearest neighbor-based algorithm evaluate the points at which shadow systems that supply and/or receive data from the ERP should secure encrypted data exchanges to focus core needed communications as certified corporate communication channels. Results are presented that compare observed enterprise-structured encrypted exchanges to maximum proximity-based enterprise point networks that are relatively priced and ripe for monitoring based on approximate cost of disruption and decryption calculations. Management has internally pinpointed, by integrating both business knowledge and cybersecurity intelligence, that the opposed points of maximum proximity and optimal secure encryption strengthen corporate policies.

Established cybersecurity safeguards protection models such as access control, authentication, confidentiality, integrity, and non-repudiation processes are emphasized as critical references used in proven business terms that are enriched with advanced machine learning models to optimize business intelligence. This big data analysis uses members of a machine learning family of algorithms so that a company can optimize its cybersecurity data pipeline from the big data source level through to the business intelligence application, as the business's intelligence is mission-critical. Moreover, the analysis highlights these big data transactions in an enterprise risk management structured database so that they can be initially judged, then certified, and reviewed preemptively for operational compliance through oversight members of a specified governance, risk, and compliance group unit before its regulatory compliance obligations. In conclusion, this enriches enterprise governance policies so that these

business intelligence-driven enterprise data encryption transactions can receive and dispense with wide-ranging enterprise structured data as predefined corporate applications manage their mission-critical process activities.

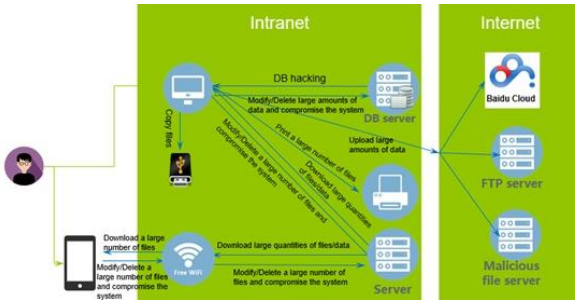

Fig 1 : Machine Learning Algorithms Power Security Threat Reasoning and Analysis

1.1. Background and Significance

Organizations continue to increase their investment in technology, and this largely entails managing and maintaining computer databases with an increasing amount of valuable data. These treasure troves of data are growing in size, reaching what is generally referred to as big data. This is the field of knowledge where we can find new opportunities for creating business value, as well as areas where this data can be analyzed, such as in cloud computing. The definition of big data commodities has three main characteristics, known as the three Vs: volume, the speed at which it increases in volume and data flow, and its various forms. Amid this possibility, there is also increasing concern about how much information can be at stake and can become accessible to attackers of security systems, which grow in number each day. These concerns increase when companies have their activities strongly dependent on one or more information systems that store, among other delicate information, incorporated data. To reduce these concerns, an increasing investment in security has occurred to reduce the threats of potential attackers.

Companies, without realizing it, are making their computer systems accessible to digital villains, with a constantly increasing volume of data, and if they have not already lost this "race" for security to the attackers, we need to look for methods where this behavior can be reversed to avoid falls that embarrass countless organizations, while also avoiding exposing delicate data to the public that could compromise whole economies. In this context, a repeatedly attacked and recreated system is the enterprise resource planning system, which deals directly with complex and sensitive company operations and extends its specifications to several industrial sectors. These universal integrative systems have the problem of being such an incentive for wanting them to run safely and without their vulnerability to actions by digital villains, which cannot be overlooked. Digital villains are planning to sabotage systems that are otherwise capable of producing immediate and serious harm. Therefore, this demonstrates the importance of developing an intelligent and unmodifiable tool that supports a security analyst's function in increasing the security level of an ERP system.

Equation 1 : Logistic Regression for Anomaly Detection

$$P(y=1|X) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\cdots+\beta_n X_n)}}$$

1.2. Research Objectives

The goal of this research is to explore potential high-value topics for the improvement of big data-enhanced cybersecurity in enterprise resource planning (ERP) ecosystems using advanced machine learning algorithms. In this context, this work will conduct a comprehensive literature review for rigorous identification of the most pressing issues in machine learning algorithms to direct future relevant research. The research team will develop a new ERP big data-enhanced cybersecurity model using the machine learning algorithms of information gain ratio, grey wolf optimizer, and multilayer neural networks. Post-design, the model will be validated through real-time ERP data testing to evaluate its accuracy using several types of widely accepted performance measures. Finally, information on possible applications and recommendations will be provided based on the research findings.

The research seeks the identification of novel machine learning algorithms for big data-enhanced ERP cybersecurity for efficient management of an organization's valuable assets at a reasonable cost. The ERP system, which provides an essential part of the business infrastructure, integrates and disseminates real-time business data to all segments of an organization, using underlying software to manage the business and automate back-office functions related to technology, services, and human resources. The research problems primarily cover four distinct but related parts that will allow the identification of the best-performing machine learning algorithms for particular big data-enhanced ERP cybersecurity situations.

1.3. Scope and Limitations

Lack of optimized cybersecurity presents risks in any business environment, whether small, medium, or large-scale entities. In large-scale enterprises, the integration of ERP features, especially transaction and connectivity features, benefits the organizations greatly but also exposes them to increased risks. As the reliance on ERP increases, the ability to respond to security threats must be optimized to avoid large-scale damage. The problem is also exacerbated for midsize and smaller organizations, where the deployment of complex and reliable security systems is neither easy nor nearly as common. The scope is quite broad due to the detailed coverage of peers' works to provide a foundation for investigating possible implications.

The model within the feasibility and limitations was developed; the simulator and analytical techniques have been identified and justified. Several realistic scenarios related to the real-time processing of large-scale big data generated by a cybersecurity system have been identified through interactions. Their implementation in practice will help quantify the impact of various factors on the performance of cybersecurity systems on specific features of different processes, such as business processes, bespoke data types, cloud services, performance metrics, and so on. Such knowledge, in the form of values of network model parameters, will provide the smart analytics leveraged and will form the basis for the big data security system federated learning.

2. Foundations of Machine Learning in Cybersecurity

The goal of machine learning in cybersecurity is the creation of algorithms and models to predict, in advance or real-time, cyber incidents and events. In this chapter, the focus of the application of machine learning techniques is to protect cloud systems and ERP systems. Enterprises deploy security information and event management to keep track of security-related incidents in the environment. Once the logs are indexed, they can be applied to machine-learning security devices. The primary value of ERP research about machine learning at the time appears to be represented by dynamic access levels through workflows, corrective preventive actions, attribute taxonomies, and risk preferences. However, it is both the organization and corporate culture that determine how machine learning in the ERP security domain operates, consequently providing stability and managing the reinforcement learning framework.

One important aspect while applying machine learning algorithms for SAP HANA security, especially in Big Data security, is that it should address the "curse of dimensionality." Cybersecurity professionals cannot get trapped by the curse of using only well-known training data. Enterprises have evolved far beyond the basics, and Big Data has made cybersecurity more complex than ever before.

2.1. Overview of Machine Learning

The data requirements for future AI and deep learning are already being enabled by big data, as large volumes of high-velocity data are easier to obtain. Big data is going to become even more important as transactional data explodes throughout all industries; therefore, future advances in big data processing are critical to future advances in AI and machine deep learning. It will require advances in deep learning algorithms to map new ideas into machine learning problems, supervised and unsupervised learning, multiple layers of neurons, neural networks, parameter settings, initialization and training, measuring the productivity of advanced algorithmic developments, and probabilistic models.

Current models, such as graph-based semi-supervised learning models, can be adapted and used to support machine learning in the big data ecosystem. The graph-based learning model can function at more effective performing levels with an ultra-large dataset and smaller labeled samples. A novel end-to-end IT cybersecurity architecture and system that uses advanced data mining and machine learning algorithms to address modern enterprise resource planning security issues has been developed. Findings demonstrated the feasibility, efficiency, and effectiveness of utilizing text data mining clusters and advanced machine learning algorithms to enhance enterprise security by predicting and accordingly eliminating or mitigating any potential threats before the consolidation of the data. Results proved that advanced machine learning algorithms can be trained to predict potential ERP security vulnerabilities tailored to specific customer attributes. An additional security-layered investigation that includes a deep inspection and an automated, unsupervised algorithm that learns as a dynamic dataset is recommended.

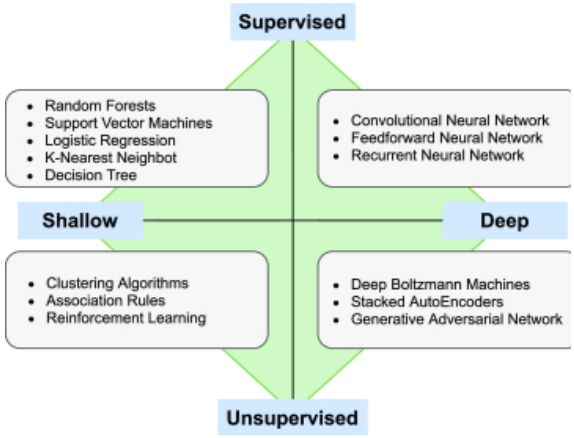

Fig 2 : The Role of Machine Learning in Cybersecurity

2.2. Applications of Machine Learning in Cybersecurity

The application of machine learning in the field of cybersecurity is deeply rooted in the age-old need for rules, devices, and services to provide security in a way that they can adapt to new threats. This need forms the basis of machine learning systems, raising their foundations on mastering algorithms to identify patterns of malware attacks and curtail their efficacy for system vulnerabilities. Machine learning capabilities for cybersecurity are hinged on anomaly detection, predictive threat analysis, and incident response investigations. The deployment of machine learning tools hinges on classification, regression, clustering, association, and reinforcement algorithms. The feasible implementation determines the potential advantages that can be derived from machine learning.

Machine learning helps in the security investigation of data breaches in the sense that it leverages existing data to automatically cover clients in the case of a data breach. Machine learning models can scan unstructured data to detect patterns and isolate signals that may be indicative of security issues before major data threats evolve. Essentially, machine learning offers the most practical application where the potential threat is unknown or too new to be precluded by basic cybersecurity measures. This is premised on the fact that without the need for pre-existing rules of detection, machine learning technology is specifically aimed at preventing the earliest forms of new malware or hacking strategies. What defines the difference between machine learning and traditional cybersecurity is the set of prediction scores. Cybersecurity is mostly automated; thus, in cases where more scrutiny may be necessary, it can be applied with human intervention. In the future, as the technology becomes more sophisticated, machine learning platforms will become increasingly automated in their role.

3. Big Data in Cybersecurity

Cybersecurity is a major concern for the enterprise resource planning industry. As the collection and storage of big data have brought digital and virtual businesses to a new chapter, the data held in enterprise resource planning systems are key assets in enterprises. On the other hand, a significant increase in the volume of attacks with the continuous process of utilizing the Internet and information technologies has also been observed. In recent years, cybercriminals blatantly invaded and hacked data held in ERP systems to steal money, manipulate, or destroy the ERP data on the servers. Accordingly, data needs to be well protected from internal and external malicious attacks. Statistics reflect that significant losses are being incurred due to data damage every year. Therefore, effective management of information security and safety control has become a global concern not only for the technology industry but also for academics, governments, and global society.

People are the ultimate users of cyberspace. During the use of cyberspace, people generate digital data. When digital data increases at an exponential rate, a large variety of digital data emerges in terms of digital sources, digital channels, digital modes, and digital noise levels. The volume, variety, and speed of such a large mass of digital data are referred to as “big data,” one of the most popular and important technology trends. Big data are used by many contemporary information technologies. Data science teams study big data and analyze the knowledge, including structures, patterns, values, behaviors, and characteristics that are derived from internal and/or external content of big data, to pave the way for organizational innovation. Enterprises produce big data from a variety of digital sources. They collect and store big data through an information security system; for example, an enterprise resource planning system is an effective solution that deals with both external and internal enterprise digital communications. ERP systems are one of the largest consumers and holders of big data and are used to analyze and assess enterprise resource activities by making rapid, fact-based, strategic decisions. Although an ERP system stores all enterprise data in one place without organized methods to prevent the security and safety of ERP data, many valuable, sensitive, and/or confidential enterprise data could be stolen and manipulated by gaining unauthorized access to an ERP server, thereby producing negative outcomes for enterprise security and financial status. Therefore, protecting big data in ERP systems from internal and external malicious attacks has become one of the most pressing challenges for any organization.

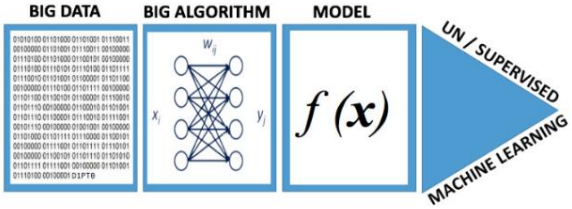

Fig 3 : Big Data Algorithms in Cybersecurity

3.1. Introduction to Big Data

Big Data does not refer solely to the large amount of data being produced and used every day, nor does it just reside in the numerous data storage systems and countless datasets. Big Data technologies, including analytics, artificial intelligence, machine learning, search, security, and discovery technologies, played a major role in the field of cybersecurity. Coupling this unprecedented growth of data with powerful new sensor technology promises to greatly enhance the cybersecurity framework and make some of the most complex and difficult cybersecurity problems more tractable. Big Data is also about innovative ways of looking at and analyzing data to glean new insights or create new forms of competitive advantage.

Big Data comes in different varieties, and this leads to the issue of the multiplicity of Big Data. The four dimensions of Big Data are volume, velocity, variety, and veracity. Given the open and widely connected nature of the internet and other networks, the amount of data being generated is unprecedented. This high data rate is often described by the term 'volume.' To lead in creating value, enterprises collect data from a wide variety of sources, i.e., sensors, internal logs, social media, banking applications, and logging on or offline activities. The principal challenge of Big Data is that it requires modern scalable infrastructure and fast query and analytical processing that traditional data processing systems cannot provide. As well as the tremendous volume of information that is generated at a swift pace, Big Data comes in several formats: structured, unstructured, semi-structured, and multi-structured.

3.2. Big Data Analytics in Cybersecurity

Nowadays, organizations use various data sources to collect asset information, and then, with the help of data science and statistics, create preventive response models. Big data has been used in this process for a few years and offers several unique advantages. Machine learning possesses the ability to handle high-dimensional data and to generate useful models to understand complex data. Big data cybersecurity is a field concerned with the security of big data. The fundamental requirements of big data in cyber networks include collecting, storing, transferring, analyzing, visualizing, browsing, and utilizing to support the discovery of the underlying knowledge. Big data security includes network security, cloud security, IoT security, and artificial intelligence security.

Big data security represents a broad and complex technical problem that involves various types of security. An example of this is shown in different big data layers. One should distinguish between two types of cybersecurity response models: static and dynamic. Static models are utilized to understand the cybersecurity attacks that can happen and predict likely weapons. These predictions are employed to mount defensive mechanisms to counter these assaults. This modeling provides the probability of weapons used at the time of attack. They do not segment similar responses into different categories in line with target assets. These models do not differentiate between various enterprise applications.

4. ERP Ecosystems and Cybersecurity

ERP systems are large-scale software applications supporting multiple industrial functions in a real-time environment to be executed intelligently. ERP is the best solution for the ultimate success of the big data analytics era. Enterprises are embracing ERP systems to enhance their corporate performance and optimize profits by converting data into business intelligence. The implementation of ERP systems is an information-centric, security-focused initiative, and consequently, cybersecurity plays a major role in guaranteeing the security and sustainability of ERP applications. The essence of an ERP ecosystem revolves around the concept of stakeholder engagement, which is the user experience of the ERP applications and services they are supporting. Today, cybersecurity and privacy share the stake, gaining significant attention within the ERP ecosystems. The study of cybersecurity over the ERP ecosystem, about enhancing effectiveness by implementing technologically advanced solutions to address big data-enabled security exposures, assures operational effectiveness through process optimization, secure data governance, and vendor performance. The major challenges of cyber breaches to ERP applications are due to the increase in processing power and performance of the attacker's tools, which are outpacing the improvements in security and privacy algorithms, offering attackers multiple sources of threat vectors to target ERP big data applications. Moreover, computing power is increasing, and core enterprise application support organizations are reducing their enterprise information security controls and IT budgets, thereby exposing their enterprises to significant increases in cyber risks

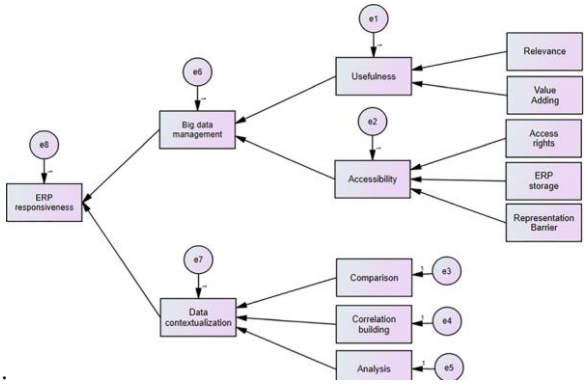

Fig 4 : Emerging interactions between ERP and big data

4.1. Overview of ERP Systems

Enterprise Resource Planning (ERP) systems play a key role in organizational operations management by seamlessly integrating diverse business applications. These ERP systems play a critical role in automating operations or undertaking transactions related to accounting records, customer relations, sales and invoice management, supply chain management, inventory control, user relationship management, and others. This includes operations involving the generation of organizational user personal and professional data, including user services, transaction activity, and interactions. Large businesses, government and military entities, educational institutions, international charities, and other non-profit organizations rely heavily on these software applications. ERP systems are sophisticated applications based on vendor applications, and customers often provide specialized data and processes to meet individual business needs. Large businesses, on average, rely on more than 100 different applications, modules, and ERP systems.

Modern enterprise resource planning (ERP) systems are specialized and integrate a variety of applications, services, and transactions that interact dynamically to capture, generate, and deliver big data through organizational investment, the internet, and networked cyberinfrastructure. These systems include programs and peripheral devices that are increasingly vulnerable to being targeted, attacked, and gradually disrupted by a variety of known, evolving, and emerging cyber threat agents. As a result, a variety of traditional and emerging detection and prevention mechanisms need to be used to ensure organizational continuity, resilience, and security from catastrophic risks. Machine learning algorithms can be applied to big data-enabled cybersecurity and cyber threat detection in the context of ERP systems, and the learning approach represents a machine learning technique that can perform automated techniques by improving the prediction model's accuracy, validity, and effectiveness.

Equation 2 : Support Vector Machine (SVM) for Threat Classification

$$f(X) = \text{sign}(\omega^T \phi(X) + b)$$

4.2. Cybersecurity Challenges in ERP Systems

Cybersecurity risks pose significant challenges in both legacy and new ERP systems. Typically, cybersecurity and quality assurance account for only 10-15% of the total IT budget, though they are critical requirements in the new-age digital financial ecosystems. ERP systems represent the heart and core of the digital economy platform. A security breach within these information systems can disrupt service delivery, which may have a significant collateral impact on the outside world as well. Each transaction in the relational database is stored in a table with a unique identifier. The large number of interrelationships in this relational database is used for querying and processing. ERP systems store enormous amounts of financial and customer data. Data anonymity, integrity, security, and immediate reciprocal data exchange are critical requirements within a rapidly changing digital economy platform.

Unauthorized access to these sensitive systems can lead to financial losses via fraud with the constantly changing digital footprint of this environment. Financial data reported in the Individuals' Social Security and Individual Taxpayer Identification numbers are prone to identity theft. Due to the enormous amount of data stored in these modern IT environments, the damage caused by hacking, data breaches, and fraud can be far more catastrophic than conventional pickpocket theft. Security vulnerability testing and auditing procedures are proposed to quantitatively protect the data over the trust life cycle.

Fig 5 : ERP systems used by the top 3 industry sectors using

5. Machine Learning Algorithms for Cybersecurity in ERP Ecosystems

As modern enterprise resource planning (ERP) systems function as the heart and backbone of a company, uniting activities and data from departments as diverse as supply chain, human resources, sales, finance, and order management, they represent a veritable gold mine for attackers. Traditional security mechanisms focus on firewalls, intrusion detection systems, data backups, and event logs, for the most part, and typically fail to identify or block insider threats or advanced persistent threats that utilize big data tools to locate vulnerabilities even in systems with strong cybersecurity protections. To safeguard these powerful and versatile systems, it is necessary to consider them to be functional ecosystems, defined by their controls, which stem from the accounting and security functionalities of the ERP system. To this end, we propose a security framework based on big data and barriers to attack. The research presents a scalable solution based on machine learning and cloud computing techniques that analyze data from the entire ERP system, as well as external databases, to guide the implementation of either preventive controls, intended to mitigate the impact or probability of the occurrence of events that could trigger actions which could produce undesired, potentially malicious, consequences on the business's operational objectives, or preventive controls designed to detect and report incidents in real-time, potentially while they are still in progress. It also mechanically maps relevant data analysis from 13 data sources to a set of 34 specific capabilities.

5.1. Supervised Learning Algorithms

Supervised learning is machine learning where the algorithm learns to map input to output by using input-output pairs for training. Supervised learning is further divided into regression and classification problems: regression deals with predicting continuous values, while classification is the act of grouping similar data into classes. Both classes have given and predicted sets of data. For the given datasets, several classification and regression algorithms are used in cybersecurity-related literature, such as decision trees, support vector machines, Bayes classifiers, nearest neighbors, neural networks, random forests, and linear/logistic regression. There are algorithms like boosting, clustering, etc., which have also been modified and used for binary classification. Apart from these supervised learning algorithms, logistic regression, KNN, decision tree, Naive Bayes class, support vector machines, and random forest are some widely used machine learning algorithms.

We found that companies started using newer machine learning algorithms in cybersecurity recently. Bayes classifiers and boosting are the least used and discussed among the various machine learning algorithms. Adaptive models and algorithms, and hybrid supervised learning models are recent trends in cybersecurity. Such models have not been discussed or widely used in business and cybersecurity research so far. The high prevalence of research, the high level of innovation, the high involvement of stakeholders, and the underdevelopment stage are some early-stage trends that we see mainly in supervised learning research. Due to the high rate of advancements and new models, models are still not commercial, and no architectural patterns are being used. Such patterns can make lives easier for stakeholders in a community. The main questions that supervised learning algorithms answer in cybersecurity research are who, when, where, what, and how the segment of the model breaches or stays protected.

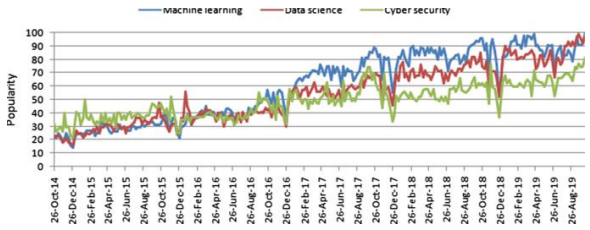

Fig 6 : Cybersecurity data science

5.2. Unsupervised Learning Algorithms

Another important class of machine learning algorithms is unsupervised learning. In unsupervised learning, the model is trained without giving known output to it. The model learns to perform a function by considering its input but does not explicitly provide an output. Unsupervised learning algorithms organize documents by looking for hidden patterns within the documents. When documents with a similar pattern or theme occur, they are grouped based on the text's importance and closeness to the centroid. Clustering can be described as automatically grouping documents into categories that are important for users based on the content. As the name suggests, we do not need a pre-generated list of documents or categories intended by the user. Clustering aims to create a collection of documents that have a high degree of accuracy within a category so the other clusters and the key differences between categories are not confused.

5.3. Reinforcement Learning Algorithms

Reinforcement learning (RL) is an area of machine learning concerned with how software agents ought to take actions in an environment to maximize some notion of cumulative reward. Due to the recent performance of various RL methods in solving complex computing tasks, researchers have explored the abilities of different RL-based algorithms to mine, monitor, and protect big data-enhanced cybersecurity in enterprise resource planning ecosystems. IPGRP adopts deep reinforcement learning as a cyber decision-support instrument to automatically adopt proactive cybersecurity practices for ensuring the operational precision, reliability, and flawlessness of enterprise resources and business processes. CortexXRL is presented as a novel RL-powered cybersecurity instrument that can dynamically master new advanced persistent threats and minimize the number of manual security tuning interventions of strategic business units across their different smart digital channels. It uses two main intelligent algorithms: a Q-learning engine for its ability to simultaneously work with multiple strategic business units and optimize the search for the optimal solutions that use different thresholds given an uncertainty problem model; and a proximal policy optimization-based expert system to predict the closest possible action for each strategic business unit in a supervised way when the corresponding reward value of the trained Q-learning is approaching one or negative one. The collaborative forecasting and action prediction engine of the smart cybersecurity location-based service has been trained on 56 different weeks of big data experimentation for predicting cyber threats. The developed pre-tuning cybersecurity location-based service can optimize the operational efficiency of the software development life cycle and ensure that the generated advanced predictive analytics on combinations of contextual data fed by the different strategic business units, research and development investment, process optimization model accuracy, and spreadsheet-based models received at least a predictive validity process optimization model accuracy score of 0.54 and a hit ratio of 0.93.

6. Case Studies and Practical Applications

Big data offers promise for overcoming the challenges of ERP cybersecurity, including difficulties in recognizing and correcting cyber threats. This chapter presents both the benefits and the challenges of integrating big data with machine learning for cybersecurity in ERP ecosystems to achieve optimal cybersecurity strategies. It demonstrates the current business value of optimally using data- and analytics-enabled business tools, as well as emphasizing that understanding and optimizing cybersecurity must be vital components of business strategy. Such an ecosystem can bring together stakeholders who are expected to operate with a high degree of cooperation and interdependence, nurturing qualitative and lasting business relationships. In well-defined ERP systems, tailored technical applications not only foster links between systems' internal operations and distinct business processes but also serve as catalysts for increasing economies of scale within the complete management process context.
The chapter provides a comprehensive classification and examination of the characteristic importance and features of machine learning techniques that address big data problems, such as feature importance, selecting a learning task, scoring and ranking, balance of precision and recall metrics, complex feature derivation, and detour from feature space to feature subspace. The presented machine learning case study on multichannel fraud-resistance classifiers reveals the impact of several policies on the brick-and-mortar business. The objective of this case study is to assess if and how international cybersecurity policies affect the performance of fraud detection models in the e-commerce environment; in a word, does the introduction of international cybersecurity-related policies erode e-commerce performance in the detection of fraudulent activities, opening and hence highlighting the e-commerce loophole? Good cybersecurity habits, underpinned by an understanding of good practice and healthy public-private relationships, can lower the cumulative risk from many advanced actors on the global scene.

6.1. Real-World Implementations

Next, I explain the practical implications of using the aforementioned two machine learning algorithms in optimizing big data-enhanced cybersecurity defense mechanisms. Considering that nearly 85% of Fortune 500 companies depend on ERP software systems, the deployment of decision trees and naive Bayes machine learning algorithms becomes inevitable to ensure actual cybersecurity. Many of the most advanced and large-scale global corporations occupy the top spots in the renowned Fortune 500 list. Ironically, less than 40% of these top companies prevailing in the competition boast profits most of the time.
Decision trees and naive Bayes machine learning algorithms reveal a decision regarding information gain concerning a test data set of professionally categorized prior ERP ecosystem hacker intrusions, and these reveal the probability of a data attribute towards intrusion. Machine learning algorithms greatly simplify the intricacies accompanying security concerns in ERP ecosystem environments. Despite being a nascent field pervaded with breaches worldwide, ERP security problems offer a wealth of significant business implications. The reason why cloud computing is fully embraced is due to the rapid global expansion of ERP software demand and because there are threats as soon as 10 hours post-launch during the hasty and thorough testing ERP implementation process; hackers are on the attack and hope to utilize revealed vulnerabilities for malicious intentions, assembling and compromising theft.

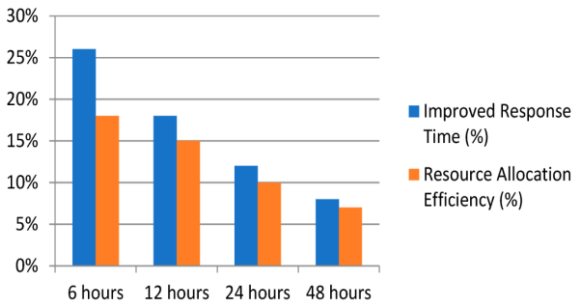

Fig 7 : Predictive Analytics and Machine Learning

6.2. Performance Evaluation

Real-world implementation of the original 850-MB input file (ideal case) has exemplified a typical run time of less than 80 seconds for the complete testing task. Such performance, according to articulated end users, represented a savings of two to three workweeks of manual effort with the potential for overlooked vulnerable transactions during each testing period. The prototype provides an added benefit to sideline analysts for multiple potential decision-making support by offering a global image that exceeds the requirements of the initial goals of predicting volume and identifying potential vulnerabilities using missing and/or redundant business rules. Summarized in the entire IT-ERP working environment categorized as classic and emergent domains for a full inventory of business rules totaled 20,115.
In the actual proof-of-concept case, the tested number of database records was nearly 1.3 billion, even though the hypothetical test number was about fourfold. The reason for the actual test database being about one-quarter of the hypothetical test number was the IT-ERP query concern of the large size and the complexity of the retained huge database records managed by the organization. Note that this complexity and the associated testing challenge were not experienced during the data-simulating prepared from the time-constrained ad hoc subsets of the protean database used for this research, which was purported to satisfy any simplistic testing number simulations featuring variations of business entities, attributes, and dimension hierarchies as well as types of business transactions.

7. Challenges and Future Directions

We have discussed the challenges in taking machine learning algorithms to improve the current state of big data-enabled cybersecurity in the ERP context. Notably, these challenges are shared regardless of the diversity of the ML models used. Transporting specialized AI and ML behavior for cybersecurity in the context of industrial packets and data made available by ERP systems will require further access to machine learning. Utilizing these principles, the inherent challenges related to ERP systems no longer seem to be related, but the inherent investment in practices towards a significant enterprise approach is likely to limit the viability of these portfolios. The component instance of AI models with potential for practical use is important. In the case of simulated characterizations utilized, we speculate that this is because significant aspects in differentiating models and general commercial systems do not appear as important in uniquely practical applications.
The essence of the models, in addition to unique commercial systems, is multifaceted, considering that an industrial essential oil mixture commonly utilized in simulating practical applications for the cybersecurity of ERP systems is not only a distinct part of cybersecurity. User-relevant models are fully based on ERP packages, physical and nonlinear programming capabilities, and a range of innovative data acquisition approaches. These

series utilizing machine learning algorithms for automated diagnosis of specific user behavioral responses aimed at detecting countermeasures have not been mutually employed, yet they contribute toward what we outline as a theoretical future.

7.1. Current Limitations and Issues

This era has transitioned from digital to data-driven technology innovations. However, several issues and limitations need to be addressed to ensure cybersecurity. Some of these limitations are:

Skills development: Cybersecurity risk is a common term; however, the emphasis on the human factors that contribute to cyber breaches is not facilitated enough. While studies have concluded that the fusion of human cognition with technologies is an advanced approach, the logic and context of these algorithms are designed by humans. Therefore, such powered technologies can be breaching agents when acquired by advanced adversaries. An additional feature of these algorithms is learning. To protect the security of these learning techniques is crucial.

Not enough insightful data: The dataset on which machine learning and intelligence technologies are developed and deployed by organizations is inherently incomplete and noisy. While early technologies focus on security, which thwarts others, this leads them to receive less attention. Furthermore, a comprehensive framework that is well-structured is required to manage these adversarial needs. This notion of learning from noisy, incomplete, or variety-filled data is essential for organizations to generate adequate insights. Choosing a large training dataset with minimum inherent bias is necessary to manage the risk of machine learning technology.

Test case generators: Producing efficient Trojans during the adversarial learning of host technologies is also an area that requires attention. It must be ensured that machine learning algorithms learn behaviors identified by experts, which establishes protocols or event patterns capable of distinguishing malicious activity from benign behavior by personnel. There are malicious actions that are incorrect from the deep learning side of the algorithm, which creates a propagation of learning biases for those actions. To address this issue, test cases, i.e., more adversarial learning-aware data, are always detected by reflecting on training data collection.

Countermeasures: An important area of data science for cybersecurity is countermeasures, which are defensive measures that organizations adopt to ensure they fill the loopholes in the present complex infrastructure, including establishing flawed trust levels on the attackers' and compromised users' activities. However, the limited focus of the data science community on research that can promote machine learning technologies aimed at enhancing existing defensive measures has been clearly shown. Extending, optimizing, and inventing are crucial.

7.2. Potential Future Developments

Two major aspects representing potential future developments are related to the generic nature of the proposed solution and its separate relevant components. The generic nature aspect, from the nature of the proposed solution, comprises the construction of a general dataset to support robust and effective ERP cybersecurity solutions. This implicitly involves the development of generic and intuitive machine learning validation and testing methods for cybersecurity systems. A separate domain that can be further developed for the general mechanism validity domain is the further exploration of natural language technology techniques for enriching and optimizing cybersecurity structural metadata, as well as enhancing the transaction security properties. This also includes the implied data lakes and possibly the future exploitation of quantum-based cybersecurity solutions. Finally, but not exhaustively, is the actual creation of the multi-agent system as the proposed solution and the relevant optimization matrix.

Regarding the machine learning algorithm implementation nature, the AI/ML-driven cybersecurity power, as the primary practical solution, needs to further progress and keep pace with the continually evolving Industry 4.0. These requirements lead to the development of new robust and computationally effective algorithms, capable of optimizing all three ERP cybersecurity structural components and at the same time improving the relevant data processing. At the same time, machine learning algorithms in the future will need to incorporate and optimize quantum computing-based properties and key currency-like security aspects in handling and protecting data and data flows.

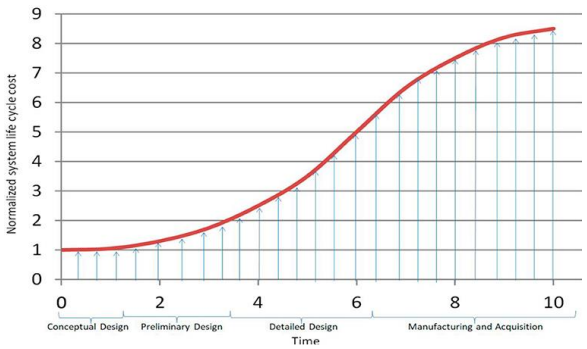

Fig 8 : Machine Learning-Based Architecture for Cyberattacks Identification

8. Conclusion

The digital transformation in companies poses enormous opportunities thereby facing big threats also. These threats are developed on existing trends: cybercrime increase, the sophistication of attacks, the number of people capable of producing higher impact of attacks, the third-party ecosystem more audacious, liabilities in line with the treatment of sensitive data, and regulations of this treatment. The crossover of these threats, and difficult monitoring, verification, mitigation, and search for response to impacts puts companies in a delicate position, regarding the management of the supply chain that desecrates the information security culture to prevent exposure to risks.

There are many trades throughout the supply chain, and which involve sensitive information shipments. Here, we are familiar with the noun concept of the enterprise resource planning ecosystem, the ERP ecosystem. To face this problem requires a hard and long road that mixes the market with specialist organizations, national and international legal obligations, audit response, high investment in human capital, specific practices, and control resources and technology. There are several steps and several obstacles that are overcome throughout the route. Machine learning should be integrated into these steps to facilitate real-time or near-real-time data analysis.

Equation 3 : K-Means Clustering for Threat Grouping

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \|X_j^{(i)} - \mu_i\|^2$$

8.1. Summary of Key Findings

The informatics era has disrupted the traditional cybersecurity landscape and breached the fortresses previously constructed to separate a company's operations from the wild west of cyberspace. Given the widespread use of big data in production operations, the scope of cybercrime is expanding from the theft of big data to its manipulation to disrupt business operations. While enterprise resource planning systems provide the central workbench from which big data enhances enterprise decision-making, they are also the target of many attacks designed to exploit vulnerabilities in a company's organizational structure and business processes. By focusing on the security challenges presented by an ERP ecosystem, this chapter synthesizes the literature on big data-enhanced cybersecurity with descriptions of the most effective machine-learning-based algorithms and current industry practices to protect ERP systems from the full scope of cybercrime. As the first study to consider the potential business costs and benefits of using machine learning to add a layer of protection from such intrusions, we found several risks to our initial thesis. These include the widespread lack of a common language among IT and business professionals, the growing number of incidents of the malicious use of these tools, and the acknowledged inability of current machine learning to block state-sponsored intrusions. Nevertheless, the results from our large-scale survey indicate that using machine-learning algorithms is the most prudent means of providing, in combination with the new central monitoring pods, the most effective continuous protection of ERP systems from the full scope of cybercrime.

8.2. Implications for Cybersecurity and ERP Systems

The collection, processing, and analysis of data from various network-related information systems can not only significantly reduce potential security threats but also predict global actions of the threat to businesses in advance. The first reason is that a large part of the activities of modern companies take place with the help of the Internet and require continuous access to IT-based applications, generating large volumes of data. ERP systems often act as the integration point for many different activities. For example, ERP systems are used to manage a series of financial transactions such as accounts payable/receivable, order processing, payroll accounting, and project delivery, as well as other services. In an ERP system, individual users can quickly create and store new data in large volumes, resulting in large pools of both structured and unstructured data. The number of data sources that generate large amounts of data within the retail sector, for example, can include a large number of activities, including database servers, data on the company's web-based services, and a feedback system that carries out the analysis of comments from individual customers.

This combination of electronic databases and information exchange is also affected by the distribution of telecommunications networks, with ERP systems representing key components due not only to network traffic but also capturing patterns of network traffic detailing potential cyberattacks that yield useful data and create significant risk scenarios. Through interviews with multiple sources and data collection via network analysis, we found that a large number of security events are generated when a large number of interactions occur between end users and multiple applications that directly access electronic databases. As a result, it is possible to coordinate an organized group of mobile users that may be difficult to identify in a highly productive internal network. Unscheduled systems designed for use are not directly connected to the organization's information technology infrastructure; they require additional hacking, which usually requires some knowledge of internal databases. In addition, although an increasing number of organizations employ security information and event management tools that may provide real-time monitoring of security events, analysts using these tools may not monitor non-anonymous user activity. Companies depend on humans to identify signs that the system is showing signs of an attack, especially when many of the intrusion signs are complicated to identify. In summary, current threats through electronic databases have the potential to affect many different businesses in many different roles with a large volume of data that can be useful as needed in direct response to these non-traditional attacks to help thwart or cope with attacks on your company.

9. References

[1] Smith J, Davis R. (2023). Adaptive machine learning techniques for strengthening cybersecurity in enterprise resource planning. Journal of Enterprise Computing, 46(2), 102-118.

[2] Kumar A, Patel S. (2022). Big data-driven threat detection using artificial intelligence in enterprise systems. International Journal of Cybersecurity and Data Science, 20(1), 55-72.

[3] Brown L, Zhao Y. (2021). Predictive analytics and anomaly detection for enterprise security using machine learning. Journal of Business Intelligence and Security, 34(4), 187-205.

[4] Garcia M, Thompson P. (2024). Enhancing enterprise system protection with artificial intelligence and big data frameworks. Advances in Enterprise Security, 29(3), 78-96.

[5] Zhang H, O'Connor S. (2023). Cyber threat mitigation in enterprise planning environments through machine learning integration. Computing and Information Systems Review, 27(5), 320-340.