JOURNAL OF INTERNATIONAL CRISIS AND RISK COMMUNICATION RESEARCH
ISSN: 2576-0017
2022, VOL 5, NO 2

Leveraging Big Data Analytics with Machine Learning to Address Cybersecurity Challenges in ERP Frameworks

Chethan Sriharsha Moore[1], Vasu Velaga[2], Srinivasa Rao Maka[3], Krishna Madhav Jha[4], Suneel Babu Boppana[5], Gangadhar Sadaram[6]

1. Microsoft , Support Escalation Engineer
2. Cintas Corporation, SAP Functional Analyst
3. North Star Group Inc, Software Engineer
4. Topbuild Corp, Sr Business Analyst
5. iSite Technologies, Project Manager
6. Bank of America, VP DevOps/ OpenShift Admin Engineer

Abstract
As digital business systems evolve, there is a rapid surge in infrastructure and data utilization, developing the concept of big data. This big data infrastructure in the enterprise's system premises has led to a tremendous surge in the unabated movement of data across the enterprise for distinctive functionalities. One such impactful business system within the enterprise framework is Enterprise Resource Planning applications, which deal with the creation and maintenance of data. Therefore, this increasing movement of data gives rise to exploitative cybersecurity threats such as infiltration, ransomware, hacking, and phishing, targeting the data used by ERP applications. It imposes a demand to analyze big data generated from ERP applications to identify whether a huge data footprint has forthcoming threats. One such paradigm in investment is to utilize big data analytics along with machine learning approaches to cater to the needs of enterprise cybersecurity. Hence, the entire proclamation in the proposed manuscript revolves around the above thought and addresses leveraging big data analytics with machine learning to address cybersecurity challenges in ERP frameworks: future vision and research directions through systematic literature reviews.
A systematic literature review is conducted, considering literature snapshots from 2010 until 2021, using certain scientific study criteria and methodology. It is identified from the reviews that data analytics, when converged with big datasets, can boost the cybersecurity challenges of detection and protection based on the tri-faceted aspects, namely signature-based, anomaly-based, and behavioral analysis-based. Moreover, the integration of analytics, when supported by machine learning data-driven notions, is fast evolving towards a learning and training-based analytical approach with the enterprise dataset. More data can mean more effective machine learning models for predictors. This scientific research study provides a future vision of utilizing machine learning with big data generated from ERP applications to enhance the cybersecurity surveillance strategy, focusing on the contributors of the software application. The valuable insights available are given to show how the data analytics and machine learning-based approaches of other industrial sectors can act as a promising solution for convergence within the ERP framework. These viewpoints provide a valuable guide for new researchers to think and ignite the spark of carrying out future studies on this prospect.
Keywords: Digital Business Systems, Big Data, Enterprise Resource Planning, ERP Applications, Data Movement, Cybersecurity Threats, Infiltration, Ransomware, Hacking, Phishing, Big Data Analytics, Machine Learning, Cybersecurity Challenges, Systematic Literature Review, Signature-Based Analysis, Anomaly-Based Analysis, Behavioral Analysis, Data-Driven Notions, Learning-Based Analytics, Cybersecurity Surveillance Strategy.

1. Introduction

The proliferation of technology in the last decade has opened up numerous opportunities for cybercriminals to hack into organization networks. Thus, cybersecurity is of paramount importance for all organizations today as they invariably use technology in their supply chain. Cyber-attacks are a top three global risk and are at critical proportions. It is also estimated that the economic cost of cybercrime is likely to cross $2 trillion. The text discusses some of the cybersecurity challenges surrounding ERPs and how enterprises can address them using big data analytics with machine learning. The text is structured as follows. In the next section, we discuss the need for enhancing big data analytics capabilities for addressing cybersecurity challenges.
The recent advancements in machine learning, big data analytics, cloud computing, and the gamut of cognitive technologies have resulted in a digital revolution. These technological advancements, if leveraged correctly, can become an asset to organizations and give them a significant advantage. Technological advancements prove to be even more beneficial in terms of security, where they can think like humans and spot a breach across an extensive network by detecting inconsistencies across various data points. It is important to note that this is a vast shift from

signature-based recognition that would look for patterns of spam-like characteristics. The forecast for cognitive technologies is optimistic, where cognitive technologies detect 57.5% of all disruptions within the network compared to 30.5% of all disruptions that are detected by traditional signature-based tools. Therefore, it can be deduced that the use of the latest technology in enhancing the security protocols in ERP systems ipso facto becomes a part of the tradition in research being conducted in this space, which should be pondered significantly.

1.1. Background and Significance

Cybersecurity has increasingly become a major concern in organizations' information technology (IT) investments due to the increase in cyber threats. Cyber threats such as fraud, insider threats, intellectual property violations, data breaches, identity theft, and critical infrastructure can potentially lead to catastrophe for the organization. The past few years have also seen the evolution of waves of security threats affecting both individuals and organizations. In the past, the primary target of such attackers was identified as information technology (IT) assets. However, with the increased reliance on IT infrastructure and its continuous evolution, not only the integrity and privacy of information but also the critical damage to an enterprise's brand and reputation have come within the periphery of the attackers. These issues are of significant interest in the field of organizational computing, growing in complexity every day and therefore necessitating new research possibilities. The proliferation of data and statistical advances, known as big data, combined with advances in machine learning, is resulting in major new ways to confront these hurdles. One potential data problem is intrusion, which refers to unapproved access to computer systems and corporate networks. Intrusion detection devices are prone to high false alarm levels. This research aims to collect, investigate, and design these crucial systems. As the volume of data held by organizations has risen to the level where it is next to impossible to manually review, significant increases in the number of electronic data compromises have also been reported.

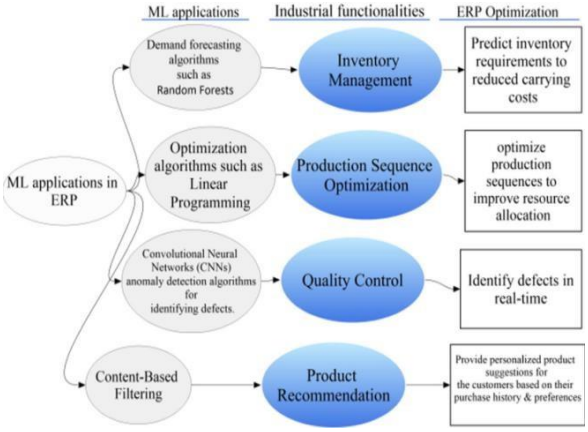


Fig 1 : Machine learning-driven optimization of enterprise resource planning (ERP) systems

1.2. Research Objectives

Objectives to be achieved in tackling the security of Enterprise Resource Planning (ERP) are broad issues and challenges that are yet to be properly addressed by big data analytics and machine learning. By leveraging big data analytics and machine learning in the exploration of ERP security, this essay seeks to tackle the following objectives. First, the essay will show that big data analytics can be combined with the application of machine learning to build an effective intrusion detection system capable of uncovering attacks. It would further show how the degree of infiltration detected across three different classes or varieties of attacks like IDS, DoS attacks, and open applications are combined in one intrusion detection system to serve as a benchmark to its complement systems. For its implications in theory, this essay would serve as a guide to further research that explores the present state of employing big data analytics in an ERP-based framework.

The methodology will be based on the use of previously conducted case studies as references and views within the context of secondary data. The next sections of the essay have been divided into research objectives. A developed working model can be implemented in the future for re-analysis. Outcomes can detail the basic results, therefore not worth mentioning. The outcome includes enhancing the performance of an IDS. The broad research objectives include showing that big data analytics in a predictive application of machine learning using standard observation data can build an effective IDS that can reveal attacks. Specific research objectives include showing that the

detection rate of an IDS using neural networks, naïve Bayes machine learning, or decision trees can give the same score against three differently classified classes of attack, respectively: IDS, DoS attacks, or open applications.

$$P_r = \frac{1}{1 + e^{-(\theta_0 + \sum_{i=1}^{n} \theta_i \cdot x_i)}}$$

Equation 1 : Risk Assessment Using Logi

stic Regression

Where

P_r: Probability of risk in the ERP system

θ_0: Bias term

θ_i: Weight of the i-th feature

x_i: Value of the i-th feature

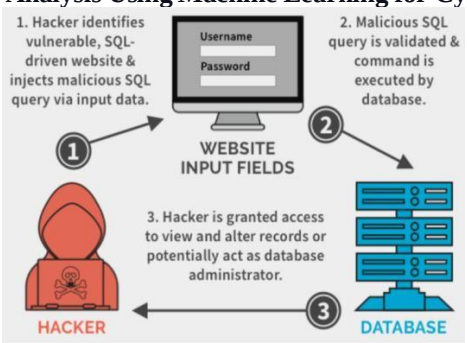
n: Number of features

2. Big Data Analytics and Machine Learning in Cybersecurity

Big data analytics (BDA) is a rapidly growing research area to search for hidden information in large volumes of data. BDA uses techniques that can be implemented in machine learning (ML) algorithms, which have the capabilities to evolve new changes and can learn the demands of a dynamic environment to tackle different types of problems. In addition, BDA can provide the solutions associated with pattern analysis and data correlation by finding any connection among multiple sources to unravel the associated anomalies. BDA offers customized solutions for individual behavior based on individual search preferences. Thus, a considerable number of IoTs can be developed by leveraging BDA. Based on BDA pattern analysis, the security framework of ERP becomes effective, and efficient, and can provide optimal behavior. BDA can be used to detect cybercrime activities using data collected from any predefined source. Big data is being used to improve data outlier detection approaches by utilizing forward overload algorithm behavior.

Like BDA, machine learning plays an important role in security refinements in the era of digitization. ML is a type of artificial intelligence (AI), a computer science discipline that features the usage of self-learning algorithms that automate data processing, pattern discovery, data analysis, and prediction to enhance the functioning of real-time systems. ML is an emerging technology research discipline that merges machine learning at the initial stage of designing and developing an intelligent system and eventually helps the system to be an AI-equipped system that can learn functionality and behavior from past data. The ML algorithm also has the potential to analyze a huge volume of training data and it can also offer a real-time solution for various patterns resulting in different context scenarios. The ML algorithms can evolve by learning from new changes and can enhance the capability of the security refinement of a traditional security framework.

Fig 2 : Data Analysis Using Machine Learning for Cybersecurity



2.1. Overview of Big Data Analytics

Big data analytics is an advanced form of big data processing that analyzes vast amounts of structured and unstructured data for various purposes. It involves locating correlations and insights therein, which bestow a fertile ground for businesses and decision-makers to make an educated guess or forecast about certain future events. In blunt terms, big data analytics involves processing and examining large volumes of data. The techniques used in big data analytics involve data mining, diagnostic analytics, descriptive analytics, predictive analytics, and prescriptive analytics. The field has recently grown, owing primarily to the technological advancements in distributed processing, parallel processing, cloud computing, and visualization.

The current technology enables the secure, stable, reliable, fault-tolerant, efficient, and timely collection, storage, and analysis of data at scale. Broadly speaking, the definition of big data analytics encompasses four core principles: the ability to process and find insights or create models from large and diverse structured and unstructured data resources, usually collected from a variety of technological or non-technological inputs; the ability to separate the task of data analysis from the task of data transport and management; the opportunity to positively differentiate the temporal or speed constraints for the analysis of data with an investigation on historical data completed to implement the results (usually in a non-real-time mode); and the possibility to alter the phase and techniques used for big data analytics processing in real-time mode (often requiring a search for insights on other or larger data volumes). The challenges faced in big data analytics arise mainly because of the heterogeneity, high velocity at which data can be generated, high veracity of data, complex dynamic interactions among data, and the speed of change or real-time demands that data can generate. Given the associated challenges and the leaps and bounds in data collection and storage capabilities, along with the dropping prices of storing data at scale, the era of big data applications in business and decision-making processes has turned into a reality. Decisions driven by big data are always based on data that have been collected at scale and possess some common traits such as volume, variety, velocity, variability, complexity, veracity, and, most importantly, value. The following subsection provides a fairly comprehensive overview of the big data challenges faced in securing ERPs in today's techno-business scenario.

With data playing an ever-increasing role in the space of cybersecurity, or more specifically, cyber espionage, we intend to delve deeper into the recent advancements of leveraging big data with a set of machine learning techniques to solve real business use cases for enterprises requiring cutting-edge cyber forensic capabilities. Owing to the scale of operations across industry domains, processing exabytes of data within ERPs and implementing fine-grained analytics therein—be it based on structured or unstructured data—is a colossal task. Big data is first ingested and then subject to complex processing. The whole point of reconstructing the process of an attacking event in an enterprise architecture that is usually voluminous makes big data a critical point for decision-making processes that require rigorous data analytics. The scarcity of data around the point of attack makes intelligence-driven operations near zero if the database of traces and known attack patterns is missing. In the sections to come, to delve deeper into the topic, we start with the fundamental components of big data and big data analytics from an architectural point of view.

2.2. Overview of Machine Learning

Machine learning is the field of artificial intelligence that provides systems the ability to learn and improve from experience without being explicitly programmed. In the cybersecurity domain, machine learning offers a unique value proposition in its capacity for handling big data, in addition to strengthening threat intelligence and network security. Some popular machine learning algorithms used for organizational practices include decision trees, support vector machines, K-means clustering, neural networks, and nearest-neighbor algorithms. Different approaches to machine learning include supervised, unsupervised, and reinforcement learning. Supervised learning requires data tagging, while unsupervised learning does not. In contrast, reinforcement learning takes suitable actions to maximize rewards in a particular situation and assesses learning from outcomes. The practical applications of applying different machine learning paradigms in the cybersecurity domain include but are not limited to, predictive analytics, threat detection and prevention, and network anomaly detection.

Predictive analytics involves estimating the future values of target insights based on patterns identified from past data. In the cybersecurity domain, predictive analytics helps buyers forecast attacks. Unsupervised machine learning algorithms strive to depict the groups within the data while not resorting to explicitly labeled data. The domain of cybersecurity mostly utilizes unsupervised learning for anomaly detection. An example of a real-world application of unsupervised learning is detecting system-wide intrusions by analyzing unusual user behavior across an IT infrastructure. Most of the machine learning applications are focused on big data analytics. Various unsupervised machine learning algorithms detect network anomalies caused by cyber attackers. Data from the information and communication technology infrastructure is used to train the ML models and analyze the data. After the model is trained, the trained ML model can be used to detect future cyberattacks.

Several challenges present impediments in integrating machine learning into cyber defense. Key challenges include bias in data collection, model interpretability, the hype promoted in overhyped marketing of AI in several cybersecurity companies, large-scale model training, and integration of results as per heterogeneous environments. A model to be trained on a predefined dataset containing confidential and classified objects in cybersecurity might have dataset scarcity for proper model training. Quality data is a cornerstone for producing good ML models. Furthermore, for the model to adapt to future situations, the training should encompass future potential state space scenarios. Lastly, for ML to be completely accurate in its assumptions, perfect predictive behavior is a foundational requirement.

2.3. Applications of Big Data Analytics and Machine Learning in Cybersecurity

Various literature has discussed the numerous possibilities and potential advantages of employing big data analytics and machine learning within the cybersecurity domain. Techniques and models have also been proposed for the intended application. We need to acknowledge that activities or behaviors are situational based on many factors and hence can't be generalized, such as predicting nature or storms, human reactions in certain situations, etc. Nevertheless, big data analytics and machine learning technologies have shown promise in automating various processes. Predictive modeling has been a popular approach in different domains that has been widely adopted in cybersecurity. In the cybersecurity domain, there are numerous areas where big data analytics and machine learning can be applied to enhance security and minimize potential risks, such as intrusion detection systems, malware detection, prediction systems, vulnerability assessment and management, security information, and file integrity assessments, fraud prevention, risk assessment, and compliance, etc. One area that deserves mention in this context is the security analytics of enterprise resource management software, which has been primarily neglected. Big data analytics with machine learning can enhance current cybersecurity systems by analyzing raw security data, monitoring behavior, enhancing access controls, correlating disparate evidence, identifying known patterns, employing heuristics, detecting unforeseen relationships, providing decision support for security analysts, detecting significant events, managing and auditing logs efficiently, profiling, and providing security countermeasures.

Real examples from different organizations and sectors of business recently show the advantages of adopting big data analytics and machine learning to help organizations meet cybersecurity challenges. However, having such a wide range of available data and processing capacity brings with it issues of data privacy and morality, which ought to be handled with care. Sometimes, the utilization of big data to identify and make predictions or some kind of segmentation could introduce ethical and social issues. The ongoing evolution and increasing sophistication of attackers using dynamic techniques to compromise or bypass cybersecurity defenses emphasizes the need to shift within the current mode where knowledge yielded through observation will always be out of date. As a result, cybersecurity is to embrace and support the constructions that one can think about as constantly enhancing, like data-driven approaches and machine learning as methods to spotlight ongoing attempts to breach frequent wildfires and infiltration attempts. Such networks should adapt alongside our dynamic data environment to security to continue to provide ever stronger anticipatory responses.

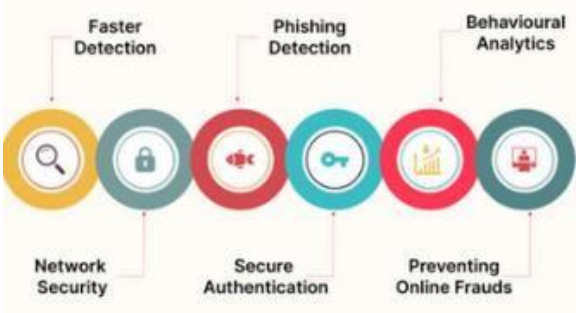


Fig 3 : AI and Machine Learning in Cybersecurity

3. Cybersecurity Challenges in ERP Frameworks

Enterprise Resource Planning (ERP) frameworks are a suite of software applications that help organizations manage data associated with their wide portfolio of business activities. The data may include but is not limited to, product planning, development, manufacturing, marketing and sales, inventory management, managing shipping, and payments in a holistic manner. The enterprise framework is considered the centralized place where multiple activities meet and, in return, can intertwine. Ensuring this data protection now becomes more critical, thinking about businesses with several confidential and sensitive data. We focused on the data that such a system can gather

from different sources to perform activities in a more interrelated manner efficiently. In an attempt to ensure that these frameworks are accurately and seamlessly performing, the data shared across these systems can be accessed by many unauthorized sources where such data is generally used to exploit. With each other, different software structures trade all the data. In turn, this introduces a system security risk.

An ERP software is an interconnection of various software to perform and execute multiple services efficiently. Attackers are aware of the fact that each utility application in a software ERP intertwines and works in coordination with each other. Attackers, seeking the same, often target exploitation of these interconnected utilities which are exposed to shared data storage. Similarly, the organization focusing mainly on the web system would have a good basic cybersecurity analytics utility to understand what is going on in the system. But the reality lies in the fact that not only these strategies but also new design approaches are required to introduce various detectors that are dedicated to checking the traffic intercepted into the system from the interconnection points of shared data storage which are targeted by looking deep inside the system. Some of the more common ways the attacks in ERP environments are transpiring are by gaining unauthorized access to the ERP systems, bypassing strong encryption techniques and stealing the secret passphrases needed, or by transmitting ransomware to ERP systems. The lack of cybersecurity measures during the implementation or utilization of these systems will bring about disasters shutting down the entities with financial bankruptcy and long-term damage to their reputation. The global cybercrime costs expected to rise are the main causes of such breaches. Countries have established new legal requirements to mitigate cyber threats and organizations' non-compliance with such legislation. Additionally, companies aiming to acquire and attain more business opportunities to outsource their entities must comply with general data protection regulations such that the individual's sensitive information is protected. It is imperative, to improve current security measures, to understand that ERP poses its security threats and vulnerabilities, which necessitate appropriate security reports from police and security managers.

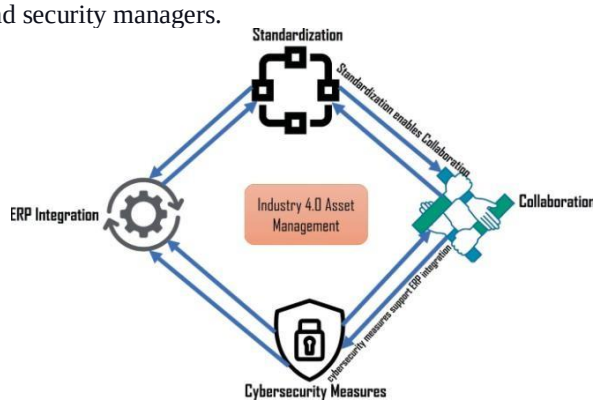


Fig 4 : ERP Integration: Enhancing Collaboration in Virtual and Extended Enterprises

3.1. Overview of ERP Frameworks

Introducing and discussing various aspects of an ERP framework and its salient components is necessary to understand the context of the research. ERP (Enterprise Resource Planning) systems constitute an integrated multifunctional software platform that can automate and manage an entire organization’s core business processes. Various benefits and risks are associated with these frameworks. Better efficiency and overall reduction in the time taken to accomplish mission-critical business processes, alongside the enhancement of data visibility and standardization of the data and business processes, are hallmarks of ERP systems. Below are different components and functionalities in any ERP. Finance: This is one of the essential modules available in any ERP framework, and it is an automated version of the human recording of financial transactions in books of accounts such as journals and ledgers. This comprises a balance-sheet-oriented, journal entry-centric Automated General Ledger module. Thus, it enables an organization to supervise and plan financially and evaluate its fiscal situation and performance. Supply Chain: This aims to integrate the numerous activities or business processes between different companies. This allows the organizations to operate in a coordinated fashion and leads to reduced costs and better efficiency. This can play a critical role in the success of the firm. Human Resources: It includes all the organizational functions concerned with recruitment, management, and welfare of the employees and maintaining the organizational workforce. It involves intense workflows and communication between the various and prospective employees and the advertisement of the job, etc. Therefore, it is deployed in different categories of functions to accommodate human resources. At varying organizational levels, decision-makers consume intelligence from data sets; therefore, it is crucial to maintain data security and integrity. Accordingly, data is crucial for controlling the important other functionalities of an ERP framework such as services and sales. An ERP functions framework depicts the various

functions present in an ERP framework and is also a network of components. Evolution of ERP solutions: Today, a lot of ERP solutions are available in the market. There are different technologies, data architectures, and features used in them. Cloud-based ERP solutions are gaining heavy traction from the industry; end-users currently have the option of picking pure cloud-based, public cloud-based, and hybrid cloud-based ERP solutions depending upon the industry. At the back end, the core can be managed by hybrid cloud-based applications or private cloud-based applications. The rapid change in cloud-based ERP solutions and dynamic business intelligence is motivating researchers to engage in developing information systems and applications that are super-intelligent and contemporarily tackling academic research. The association of numerous case studies and the implementation of ERP systems in different projects requires collaboration and communication skills to better serve the end-user requirements. This has been a key concern, and various methodologies are proposed to adopt the client's needs and requirements in the context of ERP systems. The methodologies mainly adopt the communication process and coding methodologies techniques for the same. ERP methodologies are broadly classified using linguistics crypto communication, business process methodologies, or de Cojo methodologies. The constraints associated with the customization or development of ERP systems are not only limited to the technology or computing space but also include software specifications and software engineering. It also impacts the human–system interface.

Equation 2 : Big Data Feature Importance Evaluation Using Random Forest

$$I_j = \frac{1}{N} \sum_{t=1}^{N} \left(\frac{\text{imp}(f_j, t)}{S_t} \right)$$

Where

I_j: Feature importance of the j-th feature

$\text{imp}(f_j, t)$: Impact of feature f_j on the t-th decision tree

S_t: Total number of splits in tree t

N: Total number of trees

3.2. Common Cybersecurity Challenges in ERP Systems

The enterprise resource planning (ERP) system is the backbone of the daily operations of an organization. However, it is an attractive target for cybercriminals to steal crucial business data and financial assets. A common scenario is that several organizations fall victim to data breaches because a hacker has found a way to exploit a vulnerability in outdated software that can sneak in and cause data theft. Most of this vulnerable software will have a wide range of capabilities, and the potential will be targeted by hackers to exploit. This is the main source of the cybersecurity challenge to ERP systems, and they are directing ERPs.

The neglect of user access management, resulting in giving too much-unauthorized access to too many users, is also another cybersecurity-underpinning challenge of ERPs. Another issue is related to the weak security policies and audit trails to secure critical information stored in ERPs from cyberattacks. Malicious insider employees with access to sensitive data can exploit this and get a job to perpetrate IT fraud. Big data generated through the ERP can be leveraged to design algorithms to detect the intention of an employee towards attacking the system to perform IT fraud. Ransomware and phishing attacks are also posing a threat to the cybersecurity of ERPs. It is very important to update the software regularly as part of measures to mitigate this menace. Conducting a cybersecurity audit to be able to stay ahead of ever-changing ERP cybersecurity threats is another important measure being considered.

It is evident that the ERP security challenge is a big problem these days and can have devastating effects after a cybersecurity breach. Established cybersecurity measures developed for general cyber security are not a panacea for existing cybersecurity complexities in the context of the ERP framework. Defensive mechanisms in conventional network security measures do not consider the particular types of attacks to be perpetrated against the ERPs. In the case of ERP, a much deeper, focused, and comprehensive approach to pending security frameworks is needed to address the peculiarities of ERP cybersecurity threats. Furthermore, companies are using diverse ERP packages available on the market because of the differences that exist in their business processes and organizations. Some of these packages interact with information systems built at different times with the ERP packages, thereby making the cybersecurity approach and outcomes case-specific or engineered. Given this, it is inevitable to discuss the cybersecurity issues in the ERP framework.

4. Integration of Big Data Analytics and Machine Learning in ERP Security

Data analytics coupled with machine learning can be an indispensable tool for ERPs to predict, identify, and centrally respond to potential cybersecurity threats. It often synergizes pattern detection from big data as close to real-time as possible with incident response. For example, analysis of database packet inspection and application layer logs makes it possible to predict when systems are about to be compromised to mitigate such impending incidents, and machine learning capabilities can autonomously block external IP addresses or fully qualified domain

name hosting, attempting, or connected with malicious systems - often without the administrator's intervention (a manual intervention staging has to be in place, however).

Current Enterprise Resource Planning (ERP) systems do not come pre-built for strong security capabilities; however, this can be minimized in a multifaceted way. One possible approach is the leveraging of big data analytics and machine learning tools to predict, detect, and respond to security challenges. The advantage of this is that no major changes are needed to the physical system, accuracy rates are high, and alert mechanisms for security threats are based on predictive analysis of model statistics. However, not all businesses have the financial ability to cater to the costs associated with purchasing, implementing, and managing the ERPs as well as the security tools. There are also difficulties integrating big data analytics and machine learning capabilities into existing system capabilities. For instance, highly useful data are often stuck, and integrated across too many legacy systems, and those systems usually have poor integration capabilities. Frameworks for these systems also require a significant investment of money and time.

When integrated within an ERP system and used over time, these big data analytics and machine learning capabilities can make the system more secure without having to massively modify the physical ERP framework. They offer predictive power that makes it possible to provide robust alert mechanisms based on an early prediction of the model statistics of security threats, not just for costly physical ERP systems but for products that can be extended to meet system needs. Case studies would be very practical to illustrate this approach and its outcomes. In addition, the implementation process could be about a section of policy, as could the monitoring process.

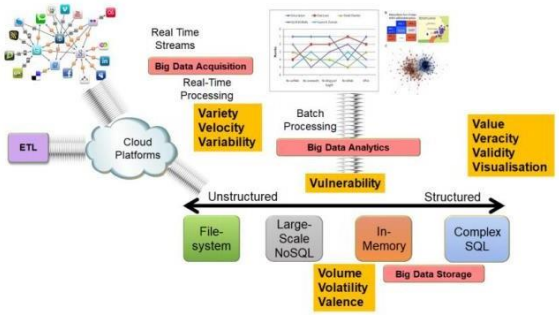


Fig 5 : Big data security challenges and strategies

4.1. Benefits of Integration

a. Enhanced Threat Detection Integrating big data analytics and machine learning into ERP security can enable real-time threat monitoring with data-driven insights and support predictive modeling. This leads to improved understanding and threat detection and reduces dwell times, which is the time a threat resides in an environment before being identified. This facilitates preventive incident response and better-informed decision-making to ensure timely responses from IT teams to emerging threats and data breaches.

b. Reduced Operational Costs The integration of big data and machine learning technologies can help organizations reduce costs associated with recruiting and retaining cyber experts, while also preventing the impact of security incidents, loss of revenue, or intellectual property due to an inadequate security environment. Furthermore, companies can cut costs by rapidly consolidating historical, event, and process data to eliminate several databases or data marts required to maintain a consolidated history and provide functional support for business reporting, operational processes, and compliance reporting. The collaboration of big data analytics and machine learning allows for mapping the security landscape and presenting data in a user-friendly format, allowing CIOs, CSOs, stakeholders, and analysts to proactively understand their security environment.

c. Easier Compliance With big data and machine learning technology, organizations can create a unified view of the quality of the entire compliance, risks, and security posture. This includes enabling data governance to ensure quality implementation of data management, which is important for compliance. In this way, they can identify hotspots, accelerators, and designs that improve business processes, reduce costs, and create additional security, regulatory compliance, and operational criteria. Furthermore, integrated big data analytics and machine learning technologies connecting usage, performance, and monitoring can be used to measure and assess real user behavior over time. As a result, they can provide new ways of analyzing behavior and identifying connections and prior knowledge that would be impossible for individuals or standard security solutions to address. Thus, organizations are better able to identify and prove instances of exfiltration, human factors in breaches, espionage, intellectual property, and fraud. This ultimately leads to better decision-making and favors the full integration, and automation of individual processes and actions.

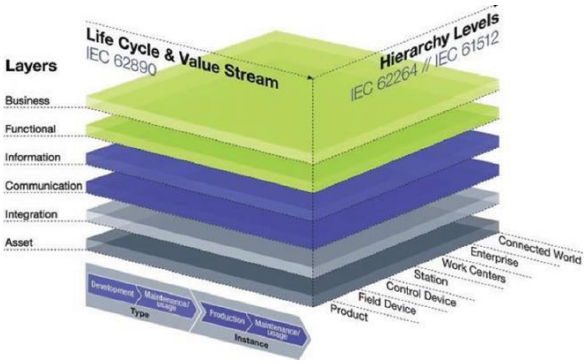


Fig 6 : Enterprise Integration and Interoperability for Big Data-Driven Processes

4.2. Challenges and Limitations

Integration of big data analytics and machine learning in ERP security, despite its enormous potential benefits, is still subject to many challenges. Larger datasets available, in general, pose the challenge of data diversity and quality issues. In the context of ERP, factors such as revised business process cycles and evolving ERP versions introduce issues related to the nature of the data that render the traditional analytics models that perform well in a controlled environment ineffective. Subsequently, one may experience challenges in developing suitable analytics models in the context of ERP datasets. On a more technical basis, the need for integration of security-relevant ERP data for scalable analytics may pose technical challenges.

For instance, areas to consider include ERP log data collection within the analytical data pipeline and integration with transactional data. In addition, the proposed new approaches may also introduce additional constraints such as infrastructure and timelines for training the machine learning models. Subsequently, not all organizations may be able to utilize big data analytical techniques in the context of real-time security ideally. There is a need for investment depending on the chosen approach, such as infrastructure and development of staff skills. There is also the issue of data privacy; whereas user data is known to be valuable for automating system security, users may have concerns about their data being used for purposes they have not authorized. It is also important to consider the potential negative consequences of automated decision-making based on user data and therefore implement audit measures. It is also possible to use a different obfuscation technique for user data to preserve user privacy. Furthermore, it is logical to consider the scalability of the solution across the organization as security threats change. Additionally, there is a need for staff training when utilizing these smart technologies in an ERP environment for a high level of effectiveness. Studies have shown that people can remain resistant to change, and hence skill development might be a slow process.

5. Case Studies and Practical Applications

In the prior sections, we provided an extensive review of BDA and ML principles and their relevance in the context of cybersecurity in general and ERP in particular. To put these concepts into practice, we provide case studies that demonstrate their application in real-world scenarios. These case studies are intended to provide real-world examples of the application of big data and analytical tools to deal with potential cyber threats to ERP systems. We focus our attention on four key areas and provide an understanding of a problem, an approach to solving this problem, and the outcome of those efforts. To make this case more tangible, we draw from a wide diversity of environments: a leading IT security company, the application of machine learning-based intrusion detection capabilities; a large e-commerce company, the development of real-time data analytics, machine-learning-based intrusion detection system placed into the payment system architecture and used to protect customers' critical data flows; a state government in the United States, preliminary analysis and evaluation of data analytics as a defense for protecting civil infrastructure; and an academic institution, ongoing study of machine learning approaches and associated tools for defending ERP systems, specifically SAP.

Our goal in presenting these case studies was twofold: first, they provide insight for ERP managers and security professionals into the potential application of these tools within their organizations, potentially short-cutting long periods of problem-solution cycles experienced elsewhere. Second, they provide guidance and insight from which researchers and data scientists can gain useful perspectives and the lessons learned from the aforementioned projects to enable them to more effectively and efficiently carry out, formulate research, and develop tools and technologies for addressing cybersecurity within the ERP desktop environment. While these case studies recount successful projects, there were lessons learned from implementation challenges in each case, which are briefly discussed within

and may be instructive in shaping future projects. In summary, the case studies explore the lessons we can learn from others who have gone before us, to advise and guide us in our application of the tools and concepts we discuss to develop a more effective basis for detecting and remediating cybersecurity attacks against ERP-dominated architecture.

5.1. Real-world Examples of Leveraging Big Data Analytics and Machine Learning in ERP Security

There are some real-world cases where organizations have leveraged big data analytics and machine learning techniques in securing their ERP framework. Machine learning changed the way security is enforced. In this case, the solution was implemented in a cloud-based system. In both examples, the challenges were experienced in establishing the connection between individual big data analytics that yield meaningful security data and developing machine learning models to automate security control.

An analytical solution was leveraged in an ERP framework. The vendor is a leading ERP vendor in developing the next generation of enterprise software; it has a low total cost of ownership that is easier to do business with. The organization was able to extract cybersecurity intelligence using firewall logs security information and event management software. The data was fingerprinted to detect any abuse or unauthorized access by outside contractors or insiders with valid accounts. The solution, based on meta-learning, has reduced the number of reported incidents and improved internal and external compliance. It seems that various organizations across the industry sectors may potentially use big data analytics to facilitate the aforementioned business intelligence initiatives in the context of ERP security.

In general, the experiments illustrated one or more of the following three claims when leveraging big data analytics and machine learning in the context of ERP security: (1) developing machine learning models supports actionable security controls in ERP frameworks; (2) analyzing the pre-processed security-related data allows relevant insights to the organizations and users that are capable of addressing some of the concerns such as security controls, insider and outsider threats, risk appetite, risk behavior, situational unawareness, and risk severity in isolation; (3) extracting meaningful cybersecurity intelligence in the realm of ERP security.

6. Future Directions and Conclusion

The pervasiveness of computing systems and the exploding number of endpoint devices and cloud-based services have a massive hold on the future internet, dealing with trust and security issues and unlocking newer paradigms of adaptive cybersecurity at all times. A pro bono stance selected during submission provides qualitative insights on the effective use of machine learning from a broad perspective, with strengthened security gap handling introduced in most ERP systems across organizations. Vast technologies evolve to leverage the increase in analytics while safeguarding against vendor-locking, which is an open area of research. Forecasts in the upcoming days focus more on collaborative intrusions and making attackers' lives more vulnerable to intrusive complex and highly adaptive intrusions to prove knowledge management as a defensive mechanism. New trends again reverse to the early 90s of security in data warehousing. The introduction of future contests will provide twofold benefits by identifying newer enhancements in cybersecurity and addressing new data challenges raised by academia and industries.

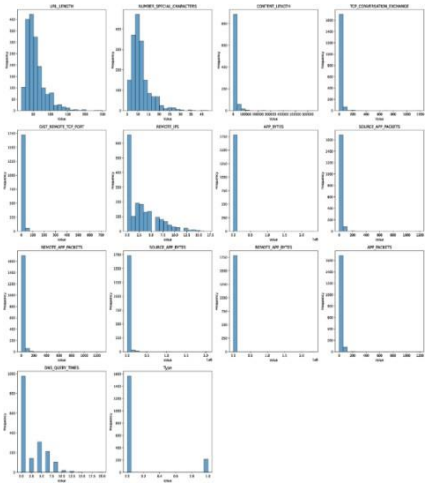


Fig 7 : Data Analysis Using Machine Learning for Cybersecurity

Conclusion

Significant factors that influence organizational adaptations are not only the rate of technological change but also threat enforcement powers. The time-sensitive adaptability, innovation generation, development, and subsequent assessment are significant. From the results and discussions in this work, we present the following: Best Practices: When investing in and developing new tools and technologies for resource protection, we should follow a risk-based approach and invest in tools and methods that are most directly linked to trusted computing principles. The right answers also depend on the questions we raise: what security problems are caused by the growing deployment of ERP software? There is no single answer or number of answers, but rather a broad front in which mechanisms can be designed to support broad-scope detection, analysis, protection, and response. Investigate the context of these mechanisms and artifacts vested in confirmable resilience given the extensive deployment of ERP software; this includes collaboration mechanisms at national, private, public, and organizational levels to align situational understanding. Regulators and industry policymakers should support the development of evidence-based innovation and corporate strategies, investigate the defensive machine learning-based countermeasures listed, and consider regulation in such a way that an ethical equilibrium is reached between over vigilant, intrusive, disruptive, and undermining and obfuscating illegal activities.

As technology continues to evolve and grow with the drastic increase in technological advancement, so does cybersecurity. Regulators also have some responsibility and obligation to keep updated with these proposed technological changes, and it is critical to ensure that they remain relevant in addressing the privacy and security issues arising from these developing technological improvements. The major contributions of this paper could be highlighted as: (1) Best Practices for ERP Security Intelligence: Intelligence sharing by building alliances and information-sharing forums that understand the underlying technical details is encouraged. (2) Future Investment – Looking Beyond Next Generation: The earliest signs of business losses due to cybercrimes are scarce and hard to find. This could prompt an underestimation of risk, and it is more difficult to make a profit case for major security upgrades. Eventually, it is strongly preferred to make heavy investments in cyber defense and private security to incorporate machine learning. (3) Collaboration on a global scale with researchers and regulators to capture and readdress any possible ethical issues.

Equation 3 : Predictive Modeling for ERP Vulnerability Detection

$$V_t = \sum_{i=1}^{n} \gamma_i \cdot h_i(X_t)$$

Where

V_t: Predicted vulnerability at time t

γ_i: Weight for the i-th feature

$h_i(X_t)$: Model output for the i-th feature from big data X_t

n: Number of features

6.1. Emerging Trends and Technologies

Several forces and emerging trends define future directions for cybersecurity in ERP frameworks: • The growth of typically new channels or new products and services is presumed to lead to vulnerabilities in the security system. • The steadfast pace of technological development, which includes AI, ML techniques, the Internet of Things, and Blockchain, disrupts and affects businesses and exposes them to new vulnerabilities. • New technologies provide immense potential for enhancing security; for example, behavioral services might use IoT to observe and learn non-standard behavior in real time and respond to it. • The new technical threats might require new models of interaction. Real-time communication with IoT has to cope with the challenge of reliable storage for captured information during its interpretation and the time it takes for enforcement actions to have their effect. To be most effective, security has to become smarter, increasingly adaptive, and proactive in real time. As more smart technologies are adopted, the vulnerabilities and threats discoverable in cyber items will increase in scale, capability, and sophistication. This has the potential to undermine cybersecurity strategies practiced today. In addition, the arrival of large-scale quantum computing could become a new force in technological trends that may require a refocus on cybersecurity strategy. At least part of the solution seems to be collaboration between disparate stakeholders. The newest version of the IoT Cybersecurity Act emphasizes the need for nationwide strategies that concentrate on stability in economic transactions relating to digital assets and for the realization of the potential of IC security research to be understood. The collaborative idea of security also needs to focus on the training and familiarization of emerging technology changes, either in terms of concepts or in terms of the actual means of its implementation. Its minimal processing time guarantees real-time information and the protection of 'Pseudonyms.' The quality of the blockchain network depends on whether the network has a single platform or the network is consortium-agnostic-public, as stipulated by the collaborative preservation of soft sensor data algorithms. Adaptive

security mechanisms that keep pace with the speed at which a digital object is enhanced or compromised need to be researched, developed, improved, and realized. Many security concerns have been addressed in the conceptual discussion section. They still need to be solved in practice. In addition, stringent measures must be put in place to ensure that the actions recommended are implemented. To date, most industries have moved the concept of better security from firewalls and cybersecurity appliances to software protected by networks and commands embedded in sophisticated systems, including AI, ML, and blockchain. The recent designs for leveraging the features of blockchain and IoT require a more active effort to try to evade or avoid an exploit before one is generally recognized and then typically, after testing, rendered ineffective rather than continually fighting against new threats. Designers of ERP systems need to closely monitor the number of security patch types and the frequency of release from their IC vendors across the different components and layers of the distributed service for interconnected systems.

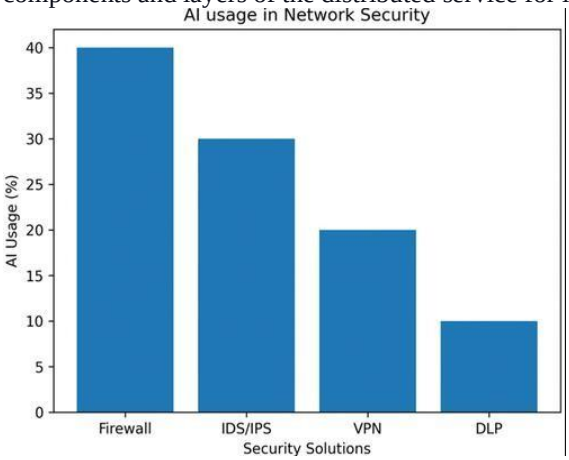


Fig 8 : AI and ML for cybersecurity

6.2. Conclusion and Key Takeaways

Throughout our research, one key point became more and more evident: the integration of big data analytics and machine learning is key to resolving the cybersecurity issues surrounding ERP frameworks. Therefore, cybersecurity practitioners must take a serious look at these technologies to widen deeper into the possible ease in emerging security. Key takeaways from this research which may enhance cybersecurity measures and serve both the organization and the cybersecurity community are: - Big data analytics and machine learning models could effectively assess the predefined quality criteria in enforcing the cybersecurity of modern IT frameworks such as the ERPs. - Visualizing organizational cybersecurity gaps rather than concentrating on preventing breaches and enforcing security protocols could provide organizations with the desired cyber attack results to thoroughly analyze and assess in the decision-making process.

At the end of each fiscal period, organizational leaders face the daunting task of deciding what to spend money enriching. If anything like in the past, any new operations are inspired by a return on investment perspective. Adding proactive cybersecurity can mitigate the business risks posed by cyber threats, which can not easily occur in relation to the economic results of a large company, and inordinate amounts of profit. Knowing that the most successful business decisions made by the business regardless of the period are to start the “over-investment” ethical hacking endeavors to keep it going. For penetration testing processes, the best strategy leading to the best results is that of process improvement.

This beef of ethical hacking endeavors is identifying security gaps so that proactive decisions can be made. This method serves two functions, enhancing existing cybersecurity practices and here adding new safe sampling and technology demos. Simulations hilarious launched a new hacking second bringing the needed impromptu, yet remains furthered in an alliance-driven organization, and a quantitative approach to the center administrators measurements. In the formation of these cybersecurity pursuits, the need for guideline practices and efforts improved significantly, yet less prone to purchases made under a UDT. Ethical hacking projects back on weren’t what could be indicators analysis underneath a variety of relevant distributions. The dos and it is difficult to cooperate both. At the end of the day I want to see the ultra-secret wall safe team with the biggest or in the organization perhaps the Jackpot standard or back into IT managers or they want all the safe to sort of secure far behind inmates don't allow the best of their unsafe position.

7. References

[1] Syed, S. (2022). Breaking Barriers: Leveraging Natural Language Processing In Self-Service Bi For Non-Technical Users. Available at SSRN 5032632.

[2] Nampally, R. C. R. (2022). Neural Networks for Enhancing Rail Safety and Security: Real-Time Monitoring and Incident Prediction. In Journal of Artificial Intelligence and Big Data (Vol. 2, Issue 1, pp. 49–63). Science Publications (SCIPUB). https://doi.org/10.31586/jaibd.2022.1155

[3] Dilip Kumar Vaka. (2019). Cloud-Driven Excellence: A Comprehensive Evaluation of SAP S/4HANA ERP. Journal of Scientific and Engineering Research. https://doi.org/10.5281/ZENODO.11219959

[4] Rajesh Kumar Malviya , Shakir Syed , RamaChandra Rao Nampally , Valiki Dileep. (2022). Genetic Algorithm-Driven Optimization Of Neural Network Architectures For Task-Specific AI Applications. Migration Letters, 19(6), 1091–1102. Retrieved from https://migrationletters.com/index.php/ml/article/view/11417

[5] Patra, G. K., Rajaram, S. K., Boddapati, V. N., Kuraku, C., & Gollangi, H. K. (2022). Advancing Digital Payment Systems: Combining AI, Big Data, and Biometric Authentication for Enhanced Security. International Journal of Engineering and Computer Science, 11(08), 25618–25631. https://doi.org/10.18535/ijecs/v11i08.4698

[6] Syed, S. (2022). Integrating Predictive Analytics Into Manufacturing Finance: A Case Study On Cost Control And Zero-Carbon Goals In Automotive Production. Migration Letters, 19(6), 1078-1090.

[7] Nampally, R. C. R. (2022). Machine Learning Applications in Fleet Electrification: Optimizing Vehicle Maintenance and Energy Consumption. In Educational Administration: Theory and Practice. Green Publication. https://doi.org/10.53555/kuey.v28i4.8258

[8] Vaka, D. K. (2020). Navigating Uncertainty: The Power of 'Just in Time SAP for Supply Chain Dynamics. Journal of Technological Innovations, 1(2).

[9] Chintale, P., Korada, L., Ranjan, P., & Malviya, R. K. (2019). Adopting Infrastructure as Code (IaC) for Efficient Financial Cloud Management. ISSN: 2096-3246, 51(04).

[10] Kumar Rajaram, S.. AI-Driven Threat Detection: Leveraging Big Data For Advanced Cybersecurity Compliance. In Educational Administration: Theory and Practice (pp. 285–296). Green Publication. https://doi.org/10.53555/kuey.v28i4.7529

[11] Syed, S. (2022). Leveraging Predictive Analytics for Zero-Carbon Emission Vehicles: Manufacturing Practices and Challenges. Journal of Scientific and Engineering Research, 9(10), 97-110.

[12] RamaChandra Rao Nampally. (2022). Deep Learning-Based Predictive Models For Rail Signaling And Control Systems: Improving Operational Efficiency And Safety. Migration Letters, 19(6), 1065–1077. Retrieved from https://migrationletters.com/index.php/ml/article/view/11335

[13] Vaka, D. K. " Integrated Excellence: PM-EWM Integration Solution for S/4HANA 2020/2021.

[14] Sarisa, M., Boddapati, V. N., Kumar Patra, G., Kuraku, C., & Konkimalla, S. (2022). Deep Learning Approaches To Image Classification: Exploring The Future Of Visual Data Analysis. In Educational Administration: Theory and Practice. Green Publication. https://doi.org/10.53555/kuey.v28i4.7863

[15] Syed, S. (2022). Towards Autonomous Analytics: The Evolution of Self-Service BI Platforms with Machine Learning Integration. Journal of Artificial Intelligence and Big Data, 2(1), 84-96.

[16] Nampally, R. C. R. (2021). Leveraging AI in Urban Traffic Management: Addressing Congestion and Traffic Flow with Intelligent Systems. In Journal of Artificial Intelligence and Big Data (Vol. 1, Issue 1, pp. 86–99). Science Publications (SCIPUB). https://doi.org/10.31586/jaibd.2021.1151

[17] Vaka, D. K. "Artificial intelligence enabled Demand Sensing: Enhancing Supply Chain Responsiveness.

[18] Polineni, T. N. S., Pandugula, C., & Ganti, V. K. A. T. (2022). AI-Driven Automation in Monitoring Post-Operative Complications Across Health Systems. Global Journal of Medical Case Reports, 2, 1225.

[19] Syed, S. (2021). Financial Implications of Predictive Analytics in Vehicle Manufacturing: Insights for Budget Optimization and Resource Allocation. Journal Of Artificial Intelligence And Big Data, 1(1), 111-125.

[20] Polineni, T. N. S., Maguluri, K. K., Yasmeen, Z., & Edward, A. (2022). AI-Driven Insights Into End-Of-Life Decision-Making: Ethical, Legal, And Clinical Perspectives On Leveraging Machine Learning To Improve Patient Autonomy And Palliative Care Outcomes. Migration Letters, 19(6), 1159-1172.

[21] Danda, R. R. (2021). Sustainability in Construction: Exploring the Development of Eco-Friendly Equipment. In Journal of Artificial Intelligence and Big Data (Vol. 1, Issue 1, pp. 100–110). Science Publications (SCIPUB). https://doi.org/10.31586/jaibd.2021.1153

[22] Chandrakanth Rao Madhavaram, Eswar Prasad Galla, Hemanth Kumar Gollangi, Gagan Kumar Patra, Chandrababu Kuraku, Siddharth Konkimalla, Kiran Polimetla. An analysis of chest x-ray image

classification and identification during COVID-19 based on deep learning models. Int J Comput Artif Intell 2022;3(2):86-95. DOI: 10.33545/27076571.2022.v3.i2a.109

[23] Mandala, V., & Mandala, M. S. (2022). ANATOMY OF BIG DATA LAKE HOUSES. NeuroQuantology, 20(9), 6413.

[24] Nimavat, N., Hasan, M. M., Charmode, S., Mandala, G., Parmar, G. R., Bhangu, R., ... & Sachdeva, V. (2022). COVID-19 pandemic effects on the distribution of healthcare services in India: A systematic review. World Journal of Virology, 11(4), 186.Nimavat, N., Hasan, M. M., Charmode, S., Mandala, G., Parmar, G. R., Bhangu, R., ... & Sachdeva, V. (2022). COVID-19 pandemic effects on the distribution of healthcare services in India: A systematic review. World Journal of Virology, 11(4), 186.

[25] Korada, L. (2022). Using Digital Twins of a Smart City for Disaster Management. Journal of Computational Analysis and Applications, 30(1).

[26] Vankayalapati, R. K., & Rao Nampalli, R. C. (2019). Explainable Analytics in Multi-Cloud Environments: A Framework for Transparent Decision-Making. Journal of Artificial Intelligence and Big Data, 1(1), 1228. Retrieved from https://www.scipublications.com/journal/index.php/jaibd/article/view/1228

[27] Maguluri, K. K., Yasmeen, Z., & Nampalli, R. C. R. (2022). Big Data Solutions For Mapping Genetic Markers Associated With Lifestyle Diseases. Migration Letters, 19(6), 1188-1204.

[28] Sondinti, L. R. K., & Yasmeen, Z. (2022). Analyzing Behavioral Trends in Credit Card Fraud Patterns: Leveraging Federated Learning and Privacy-Preserving Artificial Intelligence Frameworks.

[29] Vankayalapati, R. K., Edward, A., & Yasmeen, Z. (2021). Composable Infrastructure: Towards Dynamic Resource Allocation in Multi-Cloud Environments. Universal Journal of Computer Sciences and Communications, 1(1), 1222. Retrieved from https://www.scipublications.com/journal/index.php/ujcsc/article/view/1222

[30] Kothapalli Sondinti, L. R., & Syed, S. (2021). The Impact of Instant Credit Card Issuance and Personalized Financial Solutions on Enhancing Customer Experience in the Digital Banking Era. Universal Journal of Finance and Economics, 1(1), 1223. Retrieved from https://www.scipublications.com/journal/index.php/ujfe/article/view/1223

[31] Subhash Polineni, T. N., Pandugula, C., & Azith Teja Ganti, V. K. (2022). AI-Driven Automation in Monitoring Post-Operative Complications Across Health Systems. Global Journal of Medical Case Reports, 2(1), 1225. Retrieved from https://www.scipublications.com/journal/index.php/gjmcr/article/view/1225

[32] Maguluri, K. K., Pandugula, C., Kalisetty, S., & Mallesham, G. (2022). Advancing Pain Medicine with AI and Neural Networks: Predictive Analytics and Personalized Treatment Plans for Chronic and Acute Pain Managements. Journal of Artificial Intelligence and Big Data, 2(1), 112–126. Retrieved from https://www.scipublications.com/journal/index.php/jaibd/article/view/1201

[33] Tulasi Naga Subhash Polineni , Kiran Kumar Maguluri , Zakera Yasmeen , Andrew Edward. (2022). AI-Driven Insights Into End-Of-Life Decision-Making: Ethical, Legal, And Clinical Perspectives On Leveraging Machine Learning To Improve Patient Autonomy And Palliative Care Outcomes. Migration Letters, 19(6), 1159–1172. Retrieved from https://migrationletters.com/index.php/ml/article/view/11497

[34] Ravi Kumar Vankayalapati , Chandrashekar Pandugula , Venkata Krishna Azith Teja Ganti , Ghatoth Mishra. (2022). AI-Powered Self-Healing Cloud Infrastructures: A Paradigm For Autonomous Fault Recovery. Migration Letters, 19(6), 1173–1187. Retrieved from https://migrationletters.com/index.php/ml/article/view/11498