Kurdish Studies
December 2022
Volume: 10, No: 2, pp 731-742
ISSN: 2051-4883 (Print) | ISSN 2051-4891 (Online)
www.KurdishStudies.net

DOI: 10.53555/ks.v10i2.3759

Cloud-Based Data Integration and Machine Learning Applications in Biopharmaceutical Supply Chain Optimization

Mahesh Recharla[1*], Subrahmanyasarma Chitta[2]

[1*]Oracle EBS Onsite Lead, Biogen, recharlamaheshh@gmail.com, ORCID ID: 0009-0008-3636-4320
[2]Sr. Software Engineer, Access2Care LLC (GMR), subbuchitta@gmail.com

Abstract
Cloud-based data integration and machine learning applications are becoming increasingly popular in many industries because they provide a great deal of flexibility for managing different types of specialized data. Cloud platforms can now handle large volumes of diverse data rapidly, and they support a variety of storage services and machine learning applications for big data analytics. Creating advances in reliability, speed, and robustness on a secure platform, they permits the development of powerful enhancements to business functions. In the rigorous regulatory environment of biopharmaceutical supply chain operations, the value of cloud-based services in data integration process efficiency and capability for machine learning-based analytics is of significant importance. This text explains the use of cloud-based data integration and machine learning applications in the biopharmaceutical supply chain planning and management area, where significant but underutilized data resources exist. The use of production planning-related data can yield significant strategic and tactical advantages in reliability, responsiveness, and cost control, offering substantial commercial benefits. A secure cloud-based data integration platform can enable a high level of capability, efficiency, validation, and transparency for use in biopharmaceutical regulatory environments. The scope can increase the potential for both continuous improvement and benefits from operational analytical techniques and can enhance trust at all levels of planning and operations. Regulatory requirements for biopharmaceutical data acquisition, processing, and analysis can be addressed using a secure, high-capability platform, repository, and large-scale machine-learning application technologies. Consequently, a high-power platform methodology can be prepared for use in a variety of upgrade and value-added applications with a broad range of operational reach, which traditional internal capability cannot provide efficiently.

Keywords: Cloud-Based Data Integration, Machine Learning, Big Data Analytics, Cloud Platforms, Biopharmaceutical Supply Chain, Data Management, Regulatory Compliance, Secure Data Processing, Predictive Analytics, Business Intelligence, Operational Efficiency, Production Planning, Cost Control, Data Transparency, High-Capability Platforms, AI-Driven Decision Making, Real-Time Data Processing, Cloud Security, Process Optimization, Digital Transformation.

1. Introduction

In the pharmaceutical industry, there is an increasing demand for drug manufacturing technologies that allow for the flexible utilization of local resources, cost reduction, and the production of medicines in response to market fluctuations. However, designing and managing complex production distribution networks is still a great challenge. Here, we use cloud technologies to access data without the previously required infrastructure investment. The proposed data integration method performs data transformation and integration from private data sources to platforms in the preprocessing step, and our specialized applications then extract and apply the processed and integrated data in biopharmaceutical supply chain analysis and the optimization of integrated production-distribution networks. The use of the same database, together with supplier and customer interfaces, maximizes synergies for companies and helps develop data-sharing efforts and an innovative informational-integrated ecosystem in the industry.

Responding rapidly to market growth or decline is important for stakeholders in the supply chain; the phasing of investments and decision-making influences liquidity and operational profitability. Given the value of the drug, avoidance of waste is of key importance in the pharmaceutical industry. Each stage in the pharmaceutical supply chain has a direct impact on cost, and any inadequate management of the extended shelf life may have serious implications in the event of therapeutic need. We implemented a use case for biopharmaceutical supply chain optimization. Our model consists of a hybrid mode that integrates a mathematical model based on mixed-integer linear programming for strategic decision-making with simulation to model the intricate interactions among entities at different levels in the supply chain for better performance and scalability.

1.1. Background and Significance

Recent rapid advances in biopharmaceutical research and product development have led to increasing pressure on the biopharmaceutical manufacturing system, traditionally characterized by its small scale, high degree of process complexity, and low process reliability. These increased pressures are arising from factors that include the following: 1) the increasing number of approved biopharmaceutical products in the global marketplace; 2) the 10- to 15-year patent protection period for newly developed products that makes necessary the development of a new pipeline of commercially important biopharmaceutical

products; and 3) the transition in the focus of new product development from small molecule drugs to in vivo protein biologics. Typical biopharmaceutical products include monoclonal antibodies, therapeutic enzymes, and blood factors.
Biopharmaceutical manufacturing processes are strictly regulated by the government due to direct or indirect exposure to potentially biologically harmful materials. These processes are required to be validated by good manufacturing practices and good laboratory practices that are by regulations from government health authorities. Any deviations from these practices or any process instabilities or failures can result in significant economic or patient safety impacts. However, biopharmaceutical manufacturing processes also suffer from relatively high costs due to low throughput, low production scale, high construction and startup costs, high demand for skilled labor, limited flexibility in production, and increasing fierce global competition. In response to these increased pressures, especially those inherent in cost competition and strict process regulations, the biopharmaceutical industry is making changes concerning both facility management and product development. Many of these changes can be modeled using tools from chemical engineering. In this space, we will inject data science applications, including cloud-based data integration and machine learning technologies. Our objective is to deliver knowledge in the form of intuitively visualized multi-perspective and multi-source database information which aids in advanced process decision-making and in the cost-effective containment of time required to transform product formulation.

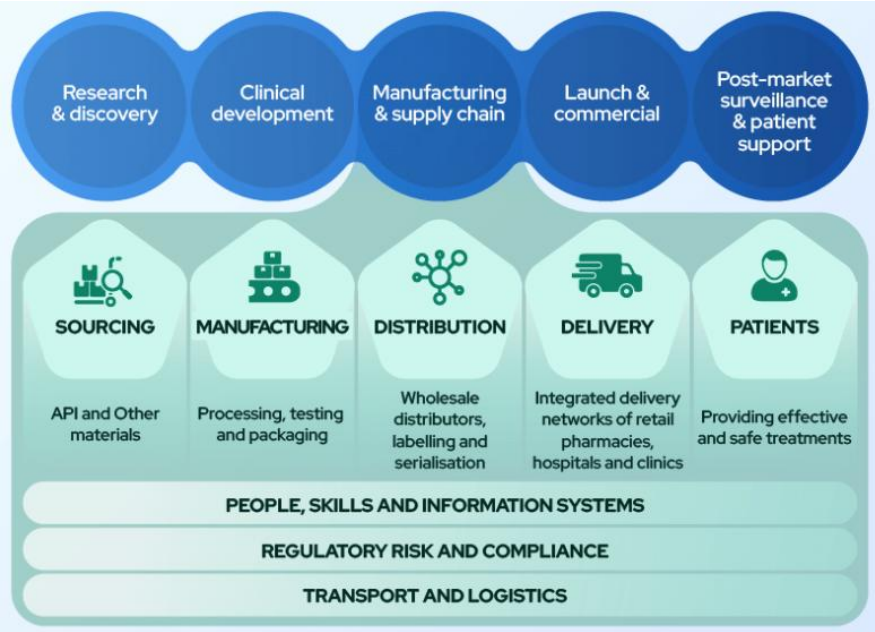

Fig 1 : Biopharmaceutical Supply Chain Management

1.2. Research Objectives

The main overall objective of this research is to optimize biopharmaceutical supply chain performance to create a more standardized, proactive, closed-loop quality management distribution infrastructure, and a more efficient, patient-centric direct-to-patient model that integrates seamlessly the provider system with patient care. Furthermore, the model reduces the amount of biotherapy needed by reducing waste and, in turn, reduces the demand for inpatient therapy, the overhead of the institution that administers the biotherapy, and the cost to public healthcare for managing these patients. This model provides the opportunity to standardize processes and outcomes for these complex patients and their providers, ultimately leading to a higher quality of life for our patients. General objectives that will achieve this outcome are to develop a multimethod approach for the optimization of a model and other quality management processes such as supply chain optimization using machine learning-based forecasting, recommendations using unused inventory forecasting, patient clinical and inventory management, patient-specific staff and facility scheduling, and location planning.

The novelty of the approach is solutions that fuse multicriteria decision-making processes required in biopharmaceutical supply chain optimization with performance- and cost-based methods under a cloud-based framework to enable a patient-specific provider model comprised of hospitals, clinics, pharmacies, specialized consultants, and patients, which creates an economic framework to increase staffing, enhance patient satisfaction, and allow the implementation of provider- and patient-directed treatment, ultimately increasing the quality of life for the patient with a proven reduced cost to the public payer of healthcare. The research results can decrease healthcare costs, stabilize treatment, improve treatment efficacy, and enhance organizations' readiness for handling biopharmaceuticals in an upcoming commercial manufacturing and treatment clinic setup.

Equation 1 : Demand Forecasting Using Time Series Model

$$\hat{D}_t = \alpha D_{t-1} + (1-\alpha)\hat{D}_{t-1}$$

$\hat{D}_t$ = Forecasted demand at time t

D_{t-1} = Actual demand at time $t-1$

$\hat{D}_{t-1}$ = Previous forecasted demand

α = Smoothing factor

2. Literature Review

Supply chain process optimization has been elaborated for many years across various disciplines such as mechanical engineering, operations research, production management, logistics management, etc. The optimization formulation can be transformed into an application domain including production scheduling, design of distribution centers, and allocation of vehicles for inbound and outbound transportation carriers, etc. Advanced technologies have been utilized in the application domain for improvement. Especially, cloud computing is being recognized as the most important infrastructure to be embedded into modern manufacturing systems, offering the ability to share resources and services over the internet at low cost. This can be considered as a model for process integration under the cloud, and we can solve the problem by using typical cloud tools to obtain visualized analysis results. Also, the cloud-based machine learning model predicts and optimizes to make it more flexible based on actual demand and supply.

The model operation of the supply chain has been studied for a long time, and the research has been extended to multi-echelon, multi-product supply chains. Supply chain performance can be improved by optimizing operational decisions, especially in biopharmaceutical SCM where product lead time is large. Balancing inventory levels with a cost service level has been examined beneficially. However, many researchers have recently realized a new criterion: optimization of the effective use of available information, and decision-making integrating modernized information technology into supply chain operations. Dependency on supplier quality assurance can be reduced when sharing some advanced product-related information like quality control technology. Demand information could be integrated with the latest information from marketing, then utilized as decision support and optimized just-in-time production plans. However, the previous literature focused on two or more systems based on separate or the purpose of visualization. This paper instead proposes a combined framework, a deep learning-enhanced cloud-based model through the combination of cloud and cloud machine learning.

2.1. Cloud-Based Data Integration in Supply Chain Management

In supply chain management, where cloud power is increasingly adopted, organizations aim for deep integration of information and capabilities both within organizational boundaries and across extended supply chain ecosystems. A digitally integrated supply chain—both internally within the firm and externally across supply chain partners—becomes a source of advantage. Information sharing and supply chain visibility are fundamental to the progress of supply chain management and have become core elements in a supply chain management strategy. The traditional approach, where the organization owns or rents hardware and licenses software or uses externally hosted EDI, requires skills for agreement, exchange, implementation, and support by the buyer or seller of information, but cannot offer the full benefit of emerging big data analytics. Cloud technology further enables data exchange selected at low cost and for deep analytic integration, capabilities that were not previously possible at such a low cost. Desirable aspects of B2B data integration and analytics include the ability to detect and respond to changing events that are relevant to the business at the velocity they arise, ideally running automated algorithms that report when pre-configured thresholds are reached or when correlations to other trigger events are observed.

2.2. Machine Learning Applications in Biopharmaceuticals

While machine learning science applied in protein design, drug discovery, drug manufacturing, and clinical data analytics in the biopharmaceutical domain is still much behind the corresponding applications in other domains, modern ML techniques are useful for a range of tasks such as protein folding, genetic pathway elucidation, splice site identification, drug discovery, and gene expression regulated by microRNA binding. Mining gene expression data in large-scale biomedical datasets has been the first machine learning application to biopharmaceutical data. It is now common to analyze data originating from biological experiments using statistical techniques and interpret results manually from the statistical analysis of the data in the biopharmaceutical field.

Characteristic expression patterns of different profiles are usually identified and analyzed. With the rapid increase of biological data, the mining process can become time-consuming as well as biased since manual interference is usually needed. Unsupervised neural networks have been used to search for density estimation of the gene expression data and combine different numerical optimization techniques to unveil those gene expression profiles during cancer intervention. Unsupervised neural networks have been used to search for the links between gene expression and drug efficiency. These networks were also combined with hierarchical clustering to analyze and classify in vivo expression data characterizing gene regulation to suggest potential research strategies and drug design. A Bayesian nonlinear PCA model for the unsupervised learning of the causal structure of time series data has been presented. Mutual information, a concept from information theory, was also used in di-predictive feature subset selection and learning non-parametric models to simulate the causal structure of gene expression. Data mining techniques including K-means and consensus clustering have been used by different research groups to mine gene expression data and address array data-related issues. The results received from these data mining techniques

and the following biopharmaceutical studies indicate that the application of these techniques could help scientists address these issues to a certain extent and improve the design and implementation of further experimental design effectively.

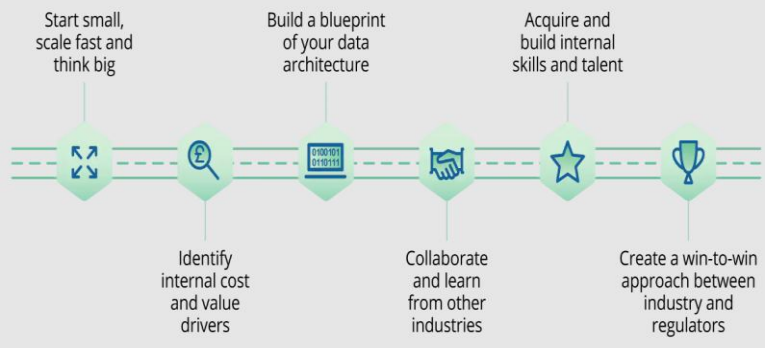

Fig 2 : AI in Biopharma Supply Chain

2.3. Integration of Cloud and Machine Learning in Supply Chain Optimization

Cloud-based data integration and BI platform driven by extensive use of ETL services is the primary underpinning technology behind any cloud-based biopharmaceutical supply chain optimization model that enjoys continuous updates of the model, frequent access, and updates of data regardless of storage location, swift and on-demand calculation, and capacity to accommodate large-scale calculations originating from source systems located anywhere in the world. The application of such technology not just in analyzing but also in guiding and shaping the decision-making process of the biopharmaceutical supply chain will guarantee the quality of cost-cutting and help to achieve far better outcomes sooner than ever before. Machine learning is a part of data science involving the development of computational algorithms possessing the ability to learn from successes and failures through pattern recognition, behavior detection, and predictions. The capability of prediction, well in advance, of disruptive events in a biopharmaceutical supply chain that may not yet have occurred is of paramount importance in arriving at well-thought-out proactive decisions and in converting supply chain constraints to strategic advantages. In the biopharmaceutical supply chain, there are many potential machine learning problems that people have worked on, but for business, most can be categorized using only three tasks: prediction, clustering or association, and anomaly detection. To develop an accurate machine learning model with state-of-the-art performance, the success of the application relies heavily on the manipulation of a massive volume of multicollinear, multivariate, highly correlated, real-world supply chain datasets.

3. Methodology

Methodology

In this paper, we consider the supply chain management of a large biopharmaceutical company and propose a model to integrate multiple data sources in the cloud. Biopharmaceutical supply chain optimization is a complex task and involves various business units. Several sources of data exist today that could potentially be used to make the supply chain run better. A biopharmaceutical supply chain involves so many different departments that it is hard for a team of even several people to understand all of the intertwined data sources and interdependencies.

The model we propose attempts to address and resolve significant problems: multiple parsing, matching, consolidation errors, and continuum of data collection across heterogeneous sources and realizing that no data source captures all of the needed information. The use of a variety of proprietary business data of the company to conduct this study allows the model to be better calibrated and more generalizable across companies. The data warehousing and rights management issues are well under control concerning very sensitive information. The methodology first dipped a few toes in cloud computing from video shooting and file movement to analysis, model building, and initial visualization of results. After some trial and error and gaining confidence, the methodology plunged much further into cloud computing waters. Time-lapse photography was used to slice the whole analysis into ever shorter intervals of cloud computational time. The conventional data processing that traditionally has been done only in loops—running code by batch mode on the local workstation, moving some intermediary results to large corporate servers, and running remaining analysis code on the servers—was then liberated from the laptop and completed on the cloud.

3.1. Data Collection and Sources

In the following, the proposed data collection and sources that are relevant to the context of the biopharmaceutical supply chain in the current work are presented. Such data will then be used to generate various time series and dynamic data sets that serve as both inputs and outputs for the machine learning and analytical models. The data required are grouped and coded logically to provide ease of data management and transformation processes to enable successful predictive or prescriptive modeling of the biopharmaceutical supply chain processes.

For the period from January 2018 to December 2023, weekly and monthly time series data related to the expression and purification processes of a biopharmaceutical are collected. Through partnerships, collaborations, and licensing activities, industry leaders have been generous and willing to provide such de-identified data. Specifically, data for weekly and monthly time series curves associated with infectivity, viability, volumetric productivity, and titer will be made available. Each curve contains about 250 samples, and developers have provided data in various formats. Additionally, data related to the dynamic scheduling, transport, and logistics of the intermodal biopharmaceutical processes and raw drug substances are acquired.

Moreover, data specific to quality attribute tests such as ELISA will be modeled. Requirements include specificity, selection of appropriate reference compounds, and validation of assumptions such as linearity between analytical response and spike concentration, and regularity.

3.2. Data Preprocessing Techniques

In the analysis of large databases, the proposed data preprocessing techniques aim to reduce storage space, speed up access to selected parts of the data, and decrease computational time. The operations include summarizing the data in some way by computing statistics, such as min, max, mean, median, or other percentiles for individual variables, or obtaining a visual summary, for example, by plotting boxplots or histograms that make general patterns of subsets explicit. Additionally, data preprocessing aims to identify and handle missing data and to deal with highly correlated input data. The treatment of missing data often involves setting the missing values based on the availability of related variables, extreme values, averages or medians, or estimates that are based on the results of clustering, classification, or regression techniques.

The removal of irrelevant variables from the dataset is based on the behavior of the variables and the criterion of a correlation matrix method. Without a prior correlation-based selection criterion, the purpose of defining worse data dimensions is false-negative exclusion by the constructed machine learning model, feature engineering, and building interpretable machine learning models. Since good machine learning practices support the use of "least data" and fewer data dimensions to avoid overfitting, the inclusion of the data variable is only justified if it contributes positively to achieving the main objective. It is essential to understand how to measure the effect of a specific variable on the model and how to quantify how well this effect can be estimated with the data available. Therefore, we present design choices for constructing interpretable machine learning models of experimental or observational data, which are two experiments for answering different scientific questions and enabling different calculations.

3.3. Machine Learning Models and Algorithms

1. MRP Type Determination To address the biological considerations of MRP types and products contained in the vaccine supply chain, we introduced a semi-supervised clustering algorithm for market demand data. We then classified demand data as flu season vaccine or one-time flu vaccine customer market demand through training on demand time series from different vaccine markets and forecasted this data. We also included the experimental data of flu season, one-time flu, and product seasonal vaccine candidates in the duration of supply chain operation. The experimental data that contained a vaccine type criterion represented the future demand information of NP antigen, a product of a one-time flu season, and that of the next flu season. According to the proportions of subcategories in experimental data, we conducted a novel quantitative MRP type determination method. This method made the decisive value of customer market demand into subtype division, and then made in-depth clustering calculations and obtained multiple mathematical equations to indicate the relationships of decision-making value and important model decision-making input indexes of principal decision nodal variables. This decision plan took the fluvirin seasonal influenza vaccine as an example to turn the soft criterion of demand data processing into a rigorous demand plan and to achieve a clear and concise automatic generation of supply chain transformation stage tasks as a high demand criterion. Finally, the effectiveness of the decision algorithm we proposed to design and develop transformation stage tasks on the pilot-scale production line transformation of the vaccine manufacturing enterprise was validated.

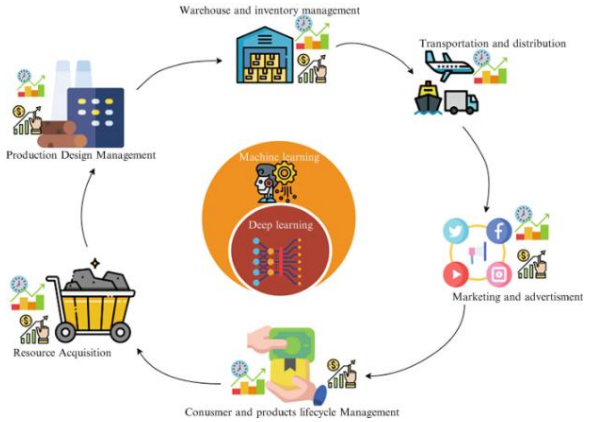

Fig 3 : Machine Learning and Supply Chain Management

2. Vaccines Production Task Scheduling We added vaccine production scheduling tasks to model the product batch creation, waiting, production, and emptying times. As the vaccine production schedule is dependent on manufacturing principles that consider batch size and type, time production principles on equipment, and time affinity processing waits and scheduling, a pre-experiment scheduling plan will be used as a decision design for the vaccine production schedule. This work brought together the equipment selection and the associated vaccine production scheduling information to create an end-to-end decision model. All of these described foundation works to support the initial MRP decisions of vaccine manufacturing tasks. High-priority influenza vaccine recommendations are linked to the functional relationship between the severity of the influenza season and the market demand for influenza. Data of this relationship are collected and analyzed, input variable relationships are determined, and training data are prepared for future machine learning prediction tasks. The ability of the model to exhibit decision task sequence results is validated.

4. Case Studies

In this chapter, we present two case studies in biopharmaceutical supply chain optimization: a data integration case and a machine learning case. The data integration case demonstrates the use of cloud-based data integration from a business user's perspective on simple data mapping. The machine learning case focuses on a predictive modeling use case and demonstrates the predictive modeling capability of business user-oriented cloud MLOps.

Biopharmaceutical supply chain optimization is a focus for both business efficiency gains and for managing dynamic global health issues. One approach to improving supply chain efficiency is by leveraging the rich data available from processes, equipment, batch reactors, analytical instruments, and material issues. This case study describes the application of cloud-based data integration techniques to collect and make sense of multiple disconnected data silos generated in the biopharmaceutical production process. We demonstrate how biopharmaceutical scientists with limited data analysis skills can use their domain understanding to derive value from the integrated data using an approach suitable for business data users. Although the purpose of the cloud-based data integration was for biopharmaceutical process understanding, the outcomes of this work also have the potential to assist with the development of predictive models to optimize the biopharmaceutical supply chain.

4.1. Real-World Examples of Cloud-Based Data Integration in Biopharmaceutical Supply Chains

Recently, cloud-based business intelligence solutions have been gaining significant popularity in all industries. These solutions offer large-scale analytics directly on external cloud data sources, from small commercial databases to public databases like market data, outsourcing partners, and social media. In the biopharmaceutical industry, companies are embracing cloud-based data integration solutions to meet the necessary regulatory obligations, optimize portfolios in discovery and development, and better develop their commercial potential. This chapter presents a selection of real-world examples that address the complex issues of data integration across the biopharma supply chain or work with large-scale data analytics that are driven by challenges in the biopharma supply chain.

Health outcomes that rely upon real-world evidence and epidemiological assessments are complementing biology and hard science in biopharma research and development activities, bridging the existing data silos. Importantly, these sources of scientific analyses and insights are effectively utilized by today's biopharma to optimize the performance of their commercial products and best address the support necessary for country-specific price negotiations and upper-tier formulary inclusion. They also contribute significantly to the stewardship of product portfolios by rationalizing the increased focus on in-licensing or out-licensing and divesting from products that no longer fit the overall business strategy. The insights offered by epidemiological assessments can also re-energize existing product portfolios and drive enhanced lifecycle management planning. Cloud-based business intelligence and analytics provide biopharma with the necessary access to existing sources of health outcomes data across a spectrum of sources, including electronic health records, clinical and patient registries, and a unique combination of claims data and life sciences through the combination of unique longitudinal databases.

Equation 2 : Supply Chain Cost Optimization

$$C = \sum_{i=1}^{n} (h_i S_i + o_i O_i + t_i T_i)$$

C = Total supply chain cost
S_i = Inventory holding cost per unit i
O_i = Ordering cost per batch i
T_i = Transportation cost per shipment i
h_i, o_i, t_i = Cost coefficients

4.2. Successful Applications of Machine Learning in Biopharmaceutical Supply Chain Optimization

The use of AI and ML in the biopharmaceutical industry has been historically limited compared to industries such as retail, banking, and manufacturing, where the technology has seen widespread adoption and a high rate of success. Given its unmet need for improved R&D and the daunting expense of bringing drugs through the approval process, the field can benefit greatly from the powerful time and cost-saving potential of AI and ML technology. This review will survey the current methods of AI and ML integration in the biopharmaceutical industry with emphasis on the biopharmaceutical supply chain, including production forecasting and scheduling.

As it stands, the majority of published ML applications in the biopharmaceutical industry are in experimental biology. Of ML applications in production, the majority are devoted to ultimately improving the performance of production processes, rather than directly acting upon the supply chain. These experimental investigations can be split into drug target identification and patient stratification. The most successful ML application areas tend to share a small set of essential properties. These place them in the reductionist camp where their fantastic predictive power does not require an accurate model motivated by physical, biological, or chemical intuition or involving complex multistep processes. In contrast, process control and optimization techniques, where applied ML has several plausible and varied applications, require such physical model-based approaches and thus form a different category. We begin with a discussion of these successful reductionist ML examples, focusing on the goals, techniques, and performances, and round off with a brief survey of the other areas.

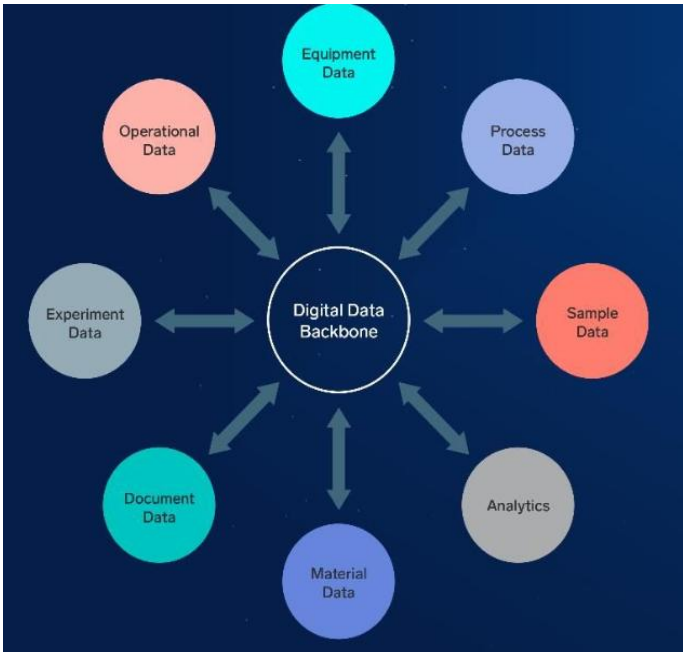

Fig 4 : Machine Learning Will Transform the Biopharmaceutical Lifecycle

5. Challenges and Opportunities

A key challenge in the context of supply chain optimization in biopharmaceuticals is the ability to effectively incorporate and take into account the details of the complex production, quality, and regulatory guidelines that are unique to specific products. There is, in essence, a tantalizing bridge to be built between highly detailed optimization models and the big data revolution in the biopharmaceutical industry, where the industry can take advantage of the insights of machine learning and big data while respecting the requirements of cGMP. While the regulatory impact assessment is a step in the right direction, enormous opportunities remain unexploited in terms of providing data and insights across products, regulatory filings, and inspections. In the context of data science, several areas of direct impact that can be addressed include decision-making needs for both proactive and reactive regulatory responses, provider and patient-specific supply chain risk assessments, and the identification and response to supply chain shocks that can emanate from geopolitical or drug security risk. The technology challenges are both innovative and leadership in nature. A significant gap exists between the expertise in the field and the availability of the software tools necessary to support robust decision-making and indicate the correct level of effort for gathering necessary data. Comprehensive and widely available predictive and learning technology platforms would strengthen regulatory decision-making by enabling regulators to anticipate where risks may be present rather than requiring agency resources to be predominantly reactive to post-event alerts.

5.1. Data Security and Privacy Concerns

Data security and privacy are major issues with cloud technology, as the cloud-based service enables a third-party server to host the data. Cloud service providers typically use a multilayered security approach that is based on physical security and proprietary network security hardware and software to protect their infrastructure. In addition, cloud service providers benefit from an economy of scale, which allows them to provide a higher level of security than many organizations could provide themselves. Yet, data protection remains a key issue for cloud-based service models due to the underlying nature of the offsite method of sharing and storing large amounts of data. There are many security and privacy concerns when adopting cloud technology for data integration and machine learning applications, which include compliance, retention, integrity, authentication, and legal issues, due to the unique challenges in an outsourcing environment where externally maintained and protected services are employed.

A comprehensive data protection and governance solution should be implemented in parallel with the cloud-based technology, ensuring the safe and effective use of cloud-based solutions while maintaining effective security control measures and data integrity. It is known that for many applications, the risks related to loss of control and validation of data are greater than the loss, theft, or corruption of the data. However, so-called 'zero access' at one end of the cloud data services spectrum, in which the cloud service provider has no access to the data at all, may limit the customer and have a significant impact on success, resilience, speed, and innovation in today's biomedical research landscape. The role of an interdisciplinary research team of cybersecurity experts and data scientists, and a range of cloud-specific resources and services, is essential to ensure the development of secure cloud-based tools and services that advance research aims and healthcare while protecting privacy.

5.2. Integration Complexity and Scalability

The ideal method for integrating data seamlessly or functionally is still an open area for research. Predefined template-based integration may be useful, but complex organizations contain lines of business that require constant discussions and negotiations simply for their fulfillment and survival. Consequently, functional integration likely involves some level of negotiation by partners. Dynamic hybrid integrations can be described by plotting multidimensional supply chain integration.

Business entanglement resulted from intersecting the static dimensions of data integration and trust integration. This suggested that customization dialogue, interoperability, and collaboration existed between these strict systems. However, different depths of entanglement require different real-time and argumentative processes. Using data integration systems as templates are not considered enough to support either a deeper or real-time entanglement, such as combinatorial electronic transactions as the preferred solution for functional integration.

5.3. Regulatory Compliance in Biopharmaceutical Industry

Regulatory compliance is one of the factors that not only introduces complexity in the integration of data but also increases the cost. Regulations in the supply chain range from data integrity to the security of data systems. The biopharmaceutical industry is regulated. The aim of streamlining good manufacturing practices, procedures, and regulations is to align with established guidelines. Legislation is responsible for consolidated traceability information, electronic means of compliance, use of electronic systems, product tracing of pharmaceuticals, transaction information, transaction history, and transaction statements. Cloud-based data integration may be at odds with this legislation because of the requirement for confidentiality. These regulations ensure data governance and traceability at every stage of the pharmaceutical supply chain.

One major disadvantage of machine learning is its complexity in decision-making. It is easier for regulatory authorities to accept simple working models rather than complex decision-making models. It might be difficult to explain the decision that has been arrived at by the more complex models, which involve a series of multiple algorithms. The gold standard for the regulatory authority in the biopharmaceutical industry is the randomized controlled experiment. The increased utilization of machine learning in the pharmaceutical industry is positively encouraged. Ideal machine learning is the cornerstone for better process development for biopharmaceutical products. It must be noted that there may be instances where model performance takes a backseat because of regulatory requirements. In this context, cloud-based data integration and machine learning require that compliance requirements be factored into system architecture. The compliance that is set in after the system has been implemented puts the life sciences and biopharmaceutical industry at increased risk.

6. Future Directions

Integrating advanced optimization, machine learning, and what-if analysis will offer scenario insight into rapid dynamic changes in an uncertain and volatile environment, which is critical for biopharmaceutical clinical drug supply chain management. This research paves the way for a digital twin for biopharmaceutical clinical drug supply with its emphasis on data processing, information storage, state estimation, feasibility, and performance-enhancing control. We believe the proposed research framework provides insights into how the technology can be developed and implemented and how the resulting automation could be understood, regulated, and optimized. Future research can explore modeling with optimization problems that are incapable of being solved through general nonlinear and nonconvex programs, where the machine learning algorithm is secured to be used to condition and transform the formulated gate problems to gain potential near-optimal solutions in a large-scale biopharmaceutical clinical drug supply chain. To illustrate the involved methodologies in addressing the biopharmaceutical clinical drug supply chain's challenging decision problems, a proper scale-up case study or joint model size is necessary and practical. It is significant to conduct pay-offs due to incorporating machine learning techniques by comparison to traditional optimization algorithms and provide additional case studies based on real-world biopharmaceutical clinical drug supply chains with existing empirical data. In the related optimization problem, robust optimization or stochastic programming can be further modeled to cope with the demand and/or supply uncertainty levels of the biopharmaceutical clinical trial supply chain. As the requirements for some biopharmaceuticals to be held in temperature-controlled environments, cold chain demand forecasts could be addressed by a controlled-growth model or bootstrap-based bias correction.

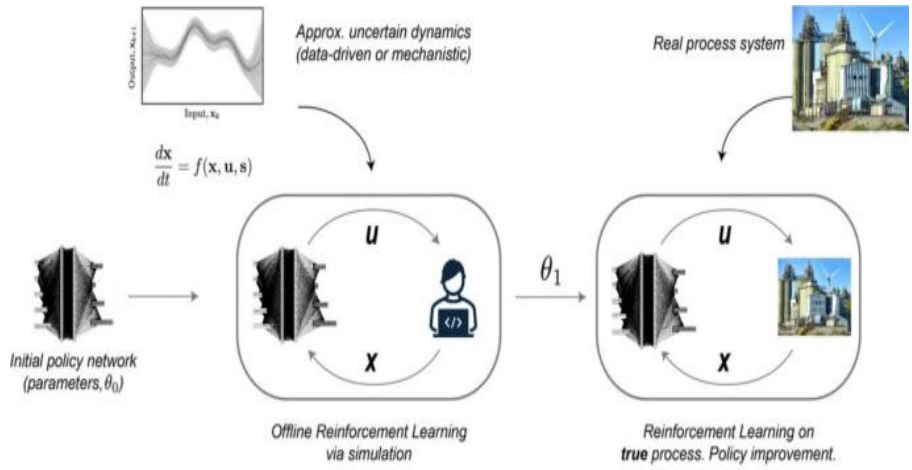

Fig 5 : machine learning from discovery to manufacturing in biopharmaceutical industry

6.1. Emerging Technologies and Trends

Emerging technologies, which are from cutting-edge research and explorations in different areas of industrial sciences, are crucial in shaping future supply chain management. This section aims to provide insight into the future direction of emerging technologies and potential trends in the aspect of supply chain management and their possible applications in the

biopharmaceutical industry. Our review discusses three main emerging technologies and predicts their trend applications in the biopharmaceutical supply chain, which are data integration, machine learning, and cloud-based computing. Applications of these three technologies in biopharmaceuticals and recent advances in these technologies are outlined. In our review, we pay attention to the factors affecting supply chain strategy derived from the implementation of these emerging technologies and propose related application trends in the biopharmaceutical industry.

Biopharmaceutical information has been growing rapidly both in size and complexity due to the diversification of business and tightening of regulatory compliance. Information size is often the consequence of multiple data sources, inherited with the inconsistency of data and structures. With the ability to capture data during almost every phase of the manufacturing process, the real problem in the biopharmaceutical industry seems to be how to re-sign information into the right recommendation or the appropriate production decision needed at the proper time. The decision-makers, who are key in all of these processes, however, need a comprehensive view of the business while considering all collected data to make more effective choices. Cloud-based technologies can help reduce these barriers. For instance, biopharmaceutical supply chains involve the business processes that support the procurement, manufacturing, and logistics of these products.

6.2. Potential Research Areas

There are opportunities to advance integrated mathematical optimization models that promote the creation of innovative global-level decision-making tools that are needed to address these strategic challenges. Such tools can account for all operational complexities across the entire supply chain that are particularly important when one aligns operational issues with customer satisfaction goals. Significant potential exists to expand the use of machine learning and data integration techniques to visualize and interpret performance data. The disciplines of machine learning, linear programming, and stochastic optimization involve training and evaluating the use of random or fixed sample elements to establish a statistically biased evaluation. As additional sources of data become available and their quality is increased, these statistical optimization tools will grow more sophisticated and comprehensive. Increasingly, advances are being made in the areas of multi-objective optimization, queueing theory, and semi-definite programming to solve the challenges related to the biopharmaceutical supply chain. However, additional research is required to further develop these methods, either individually or in combination, to enhance problem-solving capabilities across the entire biopharmaceutical supply chain. Furthermore, biopharmaceutical advances are made possible by enhanced scientific knowledge and understanding of how biological systems work at various levels of analysis. Each stage of the biopharmaceutical manufacturing supply chain thus has highly technical manufacturing requirements that occur at each stage of drug development and production. The result is a supply chain that is wasted, has poor reliability and high process variability, and responds inadequately to unexpected supply chain disturbances. The shortage of innovative and global solutions is likely due in large part to the failure of researchers to capture the true nature of biological and operational interaction in the supply chain.

Equation 3 : Machine Learning-Based Supply Chain Risk Prediction

$$P(R=1|X) = \frac{1}{1+e^{-(\beta_0+\sum_{j=1}^{n}\beta_j X_j)}}$$

$P(R=1|X)$ = Risk probability
X_j = Predictor variables
β_j = Coefficients
β_0 = Intercept

7. Conclusion

We also summarized the paper with three directions on cloud-based data integrations for pharmaceutical supply chain decisions: One is the innovation of using new data sources in traditional supply chain decisions. Two is the integration of environmental data and life cycle data with operational and financial data to support manufacturing site selection, inventory management, and transportation decisions. Three focuses on integrating blockchain with data analysis and machine learning to bring transparency and trust to organizations, to improve financial performance. These technical solutions have high values in saving cost, decreasing cycle times, improving collaboration and awareness, bringing risk management and compliance, and increasing trust, especially in an industry like biopharmaceutical and biotechnological organizations. They can potentially shorten the time to market, support product delivery during shortages and pandemics, advance life science developments, and make better medicines available for patients all over the world sooner and at a lower price. Supply chain decision-making becomes increasingly complicated as globalization and e-networking enable fast communication between different partners, with information generated and stored in more and more sources. Cloud-based data integration and machine learning can potentially unleash the power of those information silos and further advance supply chain decisions in the biopharmaceutical industry and other life science-related organizations running complex, large, and diversified supply chains. By linking the cost optimization in supply chain decisions to the value chains from the business mission of an organization, we have presented a financial-oriented decision modeling framework, with several solution categories developed by recent research across operations research, information systems, biophysical economics, digital technologies, and enterprise management communities.

7.1. Summary of Key Findings

The successful integration of cloud-based data solutions and elastic demand-driven AI can enable a more predictive biopharmaceutical manufacturing and supply chain optimization capability. Predictive AI-based models and projections can be significantly enriched from data-generating remote facilities through the employment of elastic cloud-based data services. This solution flexibility can overcome the significant challenges of on-site data-centric predictive modeling in concepts such as Digital Twin. Elastic demand-driven cloud-based data solutions can be combined into scalable predictive approaches and models that match demand variability and large data barriers, which may otherwise render internal predictive process monitoring point-of-care facilities non-economical.

AI solutions have been shown to make a key contribution to addressing the principles of biopharmaceutical manufacturing goals. These principles are founded on several major factors such as effective supplier and personal relationships, practical network optimization that links remote and near-network manufacturers, predictive actionable models for raw material supplier chemical components, levels, and supply chain performance, as well as diagnostic process performance anomaly detection. Compared to other non-GMP digital solutions, the expanded predictive capability offers several other advantages. These include knowledge-based analysis of predicted trends, early prediction of process performance deficiencies, isolation of appropriate control systems, and on-demand data-driven process predictive solution updates.

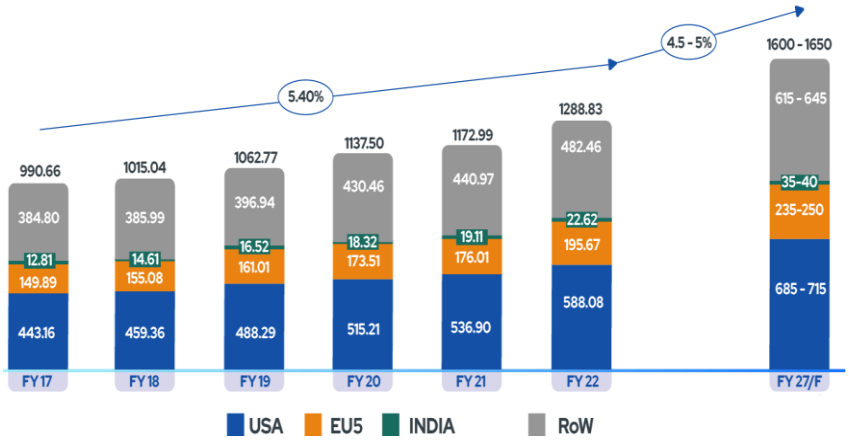

Fig 6 : Biopharmaceutical Supply Chain

7.2. Implications for Practice and Policy

The anticipated benefits unveiled by cloud-based data integration and machine learning have profound implications for the practice and policy of pharmaceutical supply chain operations and beyond. An extensive use of machine learning to support supply chain decisions can be challenging and potentially disruptive to the traditional roles of supply chain practitioners. Yet, biopharmaceutical supply chain planning, control, and continuous improvement organizations have the opportunity to seek appropriate designs and practices in redefining and refining their roles as data-driven human-plus-system supply chain decision-makers. In addition, a unique data fusion requirement of integrating process data, ERP systems, advanced manufacturing systems, and external data across the end-to-end biopharmaceutical network presents an opportunity for pharmaceutical executives and data governance organizations to define cybersecurity and data privacy policies and procedures to enable innovation, ensure patient safety, and protect intellectual property and trade secrets end-to-end.

Traditional technology enabler organization relationships, such as ERP implementation service providers, are expected to play new roles in developing and integrating data-driven analysis and machine learning models with real-time data acquisition, real-time anomaly detection, and predictive maintenance, and the specialized machine learning applications that interface with ERP or manufacturing execution systems. Furthermore, the concept of detection, control, and learning that combines anomaly detection, predictive analytics models, feedback control loops, and human-plus-algorithm collaboration can generalize to supply chain process industries beyond biopharmaceuticals where disruption has occurred or has the potential to occur, such as process manufacturing, energy, logistics, and transportation. The results strongly suggest that the practices, frameworks, and policy understandings for human-plus-algorithm decision-making – with domain expertise, risk analyses, explainable AI, and operator training within feedback loops – can have broader implications or generalizability.

8. References

[1] Balaji Adusupalli. (2022). The Impact of Regulatory Technology (RegTech) on Corporate Compliance: A Study on Automation, AI, and Blockchain in Financial Reporting. Mathematical Statistician and Engineering Applications, 71(4), 16696–16710. Retrieved from https://philstat.org/index.php/MSEA/article/view/2960

[2] Burugulla, J. K. R. (2022). The Role of Cloud Computing in Revolutionizing Business Banking Services: A Case Study on American Express's Digital Financial Ecosystem. In Kurdish Studies. Green Publication. https://doi.org/10.53555/ks.v10i2.3720

[3] Sikha, V. K. Mastering the Cloud-How Microsoft's Frameworks Shape Cloud Journeys.

[4] Challa, S. R. (2022). Optimizing Retirement Planning Strategies: A Comparative Analysis of Traditional, Roth, and Rollover IRAs in LongTerm Wealth Management. Universal Journal of Finance and Economics, 2(1), 1276. Retrieved from https://www.scipublications.com/journal/index.php/ujfe/article/view/1276

[5] Srinivasarao Paleti. (2022). Adaptive AI In Banking Compliance: Leveraging Agentic AI For Real-Time KYC Verification, Anti-Money Laundering (AML) Detection, And Regulatory Intelligence. Migration Letters, 19(6), 1253–1267. Retrieved from https://migrationletters.com/index.php/ml/article/view/11695
[6] Sikha, V. K. Building Serverless Solutions Using Cloud Services.
[7] Venkata Narasareddy Annapareddy. (2022). Innovative Aidriven Strategies For Seamless Integration Of Electric Vehicle Charging With Residential Solar Systems. Migration Letters, 19(6), 1221–1236. Retrieved from https://migrationletters.com/index.php/ml/article/view/11618
[8] Sikha, V. K. Ease of Building Omni-Channel Customer Care Services with Cloud-Based Telephony Services & AI.
[9] Kannan, S. (2022). The Role Of AI And Machine Learning In Financial Services: A Neural Networkbased Framework For Predictive Analytics And Customercentric Innovations. Migration Letters, 19(6), 1205-1220.
[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] Kishore Challa,. (2022). Generative AI-Powered Solutions for Sustainable Financial Ecosystems: A Neural Network Approach to Driving Social and Environmental Impact. Mathematical Statistician and Engineering Applications, 71(4), 16643–16661. Retrieved from https://philstat.org/index.php/MSEA/article/view/2956
[12] Ganesan, P., Sikha, V. K., & Siramgari, D. R. TRANSFORMING HUMAN SERVICES: LEVERAGING AI TO ADDRESS WORKFORCE CHALLENGES AND ENHANCE SERVICE DELIVERY.
[13] Chaitran Chakilam. (2022). Integrating Generative AI Models And Machine Learning Algorithms For Optimizing Clinical Trial Matching And Accessibility In Precision Medicine. Migration Letters, 19(S8), 1918–1933. Retrieved from https://migrationletters.com/index.php/ml/article/view/11631
[14] Koppolu, H. K. R. (2022). Advancing Customer Experience Personalization with AI-Driven Data Engineering: Leveraging Deep Learning for Real-Time Customer Interaction. In Kurdish Studies. Green Publication. https://doi.org/10.53555/ks.v10i2.3736
[15] Chakilam, C. (2022). Generative AI-Driven Frameworks for Streamlining Patient Education and Treatment Logistics in Complex Healthcare Ecosystems. In Kurdish Studies. Green Publication. https://doi.org/10.53555/ks.v10i2.3719
[16] Sikha, V. K. INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING.
[17] Venkata Bhardwaj Komaragiri. (2022). AI-Driven Maintenance Algorithms For Intelligent Network Systems: Leveraging Neural Networks To Predict And Optimize Performance In Dynamic Environments. Migration Letters, 19(S8), 1949–1964. Retrieved from https://migrationletters.com/index.php/ml/article/view/11633
[18] R. Daruvuri, "An improved AI framework for automating data analysis," World Journal of Advanced Research and Reviews, vol. 13, no. 1, pp. 863-866, 2022.
[19] Malempati, M. (2022). Machine Learning and Generative Neural Networks in Adaptive Risk Management: Pioneering Secure Financial Frameworks. In Kurdish Studies. Green Publication. https://doi.org/10.53555/ks.v10i2.3718
[20] Nuka, S. T. (2022). The Role of AI Driven Clinical Research in Medical Device Development: A Data Driven Approach to Regulatory Compliance and Quality Assurance. Global Journal of Medical Case Reports, 2(1), 1275. Retrieved from https://www.scipublications.com/journal/index.php/gjmcr/article/view/1275
[21] Karthik Chava. (2022). Redefining Pharmaceutical Distribution With AI-Infused Neural Networks: Generative AI Applications In Predictive Compliance And Operational Efficiency. Migration Letters, 19(S8), 1905–1917. Retrieved from https://migrationletters.com/index.php/ml/article/view/11630
[22] Daruvuri, R. (2022). An improved AI framework for automating data analysis. World Journal of Advanced Research and Reviews, 13(1), 863-866.
[23] Murali Malempati. (2022). AI Neural Network Architectures For Personalized Payment Systems: Exploring Machine Learning's Role In Real-Time Consumer Insights. Migration Letters, 19(S8), 1934–1948. Retrieved from https://migrationletters.com/index.php/ml/article/view/11632
[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] R. Daruvuri, "Harnessing vector databases: A comprehensive analysis of their role across industries," International Journal of Science and Research Archive, vol. 7, no. 2, pp. 703-705, 2022.
[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] Komaragiri, V. B., & Edward, A. (2022). AI-Driven Vulnerability Management and Automated Threat Mitigation. International Journal of Scientific Research and Management (IJSRM), 10(10), 981-998.
[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] Daruvuri, R. (2022). Harnessing vector databases: A comprehensive analysis of their role across industries. International Journal of Science and Research Archive, 7(2), 703-705.

[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

[35] Harish Kumar Sriram. (2022). AI Neural Networks In Credit Risk Assessment: Redefining Consumer Credit Monitoring And Fraud Protection Through Generative AI Techniques. Migration Letters, 19(6), 1237–1252. Retrieved from https://migrationletters.com/index.php/ml/article/view/11619

[36] Shashi Thota, Subrahmanyasarma Chitta, Venkat Rama Raju Alluri, Vinay Kumar Reddy Vangoor, & Chetan Sasidhar Ravi. (2022). MLOps: Streamlining Machine Learning Model Deployment in Production. African Journal of Artificial Intelligence and Sustainable Development, 2(2), 186-205. https://ajaisd.org/index.php/publication/article/view/37

[37] Dheeraj Kumar Dukhiram Pal, Jenie London, Ajay Aakula, & Subrahmanyasarma Chitta. (2022). Implementing TOGAF for Large-Scale Healthcare Systems Integration. Internet of Things and Edge Computing Journal, 2(1), 55-101. https://tsbpublisher.org/iotecj/article/view/77

[38] Pranadeep Katari, Shashi Thota, Subrahmanyasarma Chitta, Ashok Kumar Pamidi Venkata, & Tanzeem Ahmad. (2021). Remote Project Management: Best Practices for Distributed Teams in the Post-Pandemic Era. Australian Journal of Machine Learning Research & Applications, 1(2), 145-166. https://ajmlra.org/index.php/publication/article/view/82