Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 4 Issue 3 (2024)

Optimizing Enterprise ERP and E-Commerce System Integration with Machine Learning

Dr Alkawati Magadum,
Assistant Professor, MIT-ADT University College of Management, Loni Kalbor, Pune, India.
alkawati.magadum@mituniversity.edu.in

Dr Monica Goud,
JSPMS University, Wagoli, Pune, Maharashtra, India. gaudmonica@gmail.com

Dr Rachana Vasant Chavan,
Assistant Professor, Pune Cambridge Institute of Management and Computer Applications, Ambegaon, Pune, India.
rachana21.shinde@gmail.com

Abstract –The national economy and continuous modernization have been greatly impacted by the fast growth of e-commerce. Because of distinctive product features including high death and damage rates, the plant sector stands out among e-commerce sales model industries. Because of the industry's more stringent e-commerce and logistical standards, online plant stores may find it exceedingly challenging to operate. A large body of research has pointed to the importance of knowing the current state of a product's market and the specifics involved, including their size, purpose, and kind. Based on a case study, this report suggests an optimization method to address the issue of surplus warehouse storage in online plant stores.

***Keywords*—**Gradient Boosting Machines (GBM), Support Vector Machines (SVM), E-Commerce, Enterprise resource planning (ERP).

I. INTRODUCTION

Due to the rapid expansion of the information industry and the extensive usage of the Internet on a global scale, e-commerce and Internet of Things (IoT) technologies has a solid foundation. Modern e-commerce is defined by its virtuality, individuality, and portability. A system that facilitates the scheduling and online sale of products is one example of an e-commerce platform[1]. Therefore, information technology is essential for e-commerce platform development, but a solid supply chain network system is also necessary to accommodate the environment and fulfill the expectations of e-commerce platform growth. Enterprise resource planning (ERP) is a new way for businesses to consolidate their data and operations, made possible by developments in information technology. ERP stores data related to various departments and jobs in the firm in a single database. ERP is now used by almost every type of company, no matter how big or little. The main goal of implementing ERP is to centralize and simplify all of the company's data and processes, making them more accessible and enhancing their efficiency. One common way ERP systems accomplish this is by combining several pieces of software into one database[2]. ERP is a method for managing a business's money, employees, and other assets in an effective and efficient manner. It involves directing several functional areas and those pertaining such as sourcing, manufacturing, advertising, and the following: logistics, quality, finances, human resources, strategy maintenance, and operations. ERP software implementation comprises more than just installing the software; it contains involves defining the processes that a business must adhere to; this comprises figuring[3]. Various departments, and even other departments altogether, are involved in the procedure, which places a burden on the different levels of permission, as well as the checks and restrictions implemented at different stages it is necessary to make certain tough and essential decisions in order to implement them effectively across all industries. This study set out to offer a practical solution that would address all of the needs of small businesses so that they could stay afloat and expand their operations in the face of heavy competition from other service providers. Now more than ever before, even the smallest businesses have the opportunity to have an online presence, thanks to the proliferation of internet users. It is often believed that all small and medium-sized firms (SMEs) use information technology (IT) similarly, regardless of their size. Several credible scientific sources have agreed that commerce encompasses all types of trading conducted with the aim of generating a profit[4]. Trade and other types of economic activity began around the same time as humans first appeared on the scene and began to develop as a species. To meet basic needs, these exchanges merely involved exchanging goods in the beginning. It has changed through time as a result of both economic and technological developments. It is clear that various groups have used varied business practices from prehistoric times to the present. In today's generally acknowledged information society, products and services are introduced to consumers through television and online sites. The sales make it easy for people to buy more than they actually need. Individuals' consumption habits intensify and consuming societies are developed. People may just throw away products that were painstakingly crafted. Online shopping helps keep prices low by cutting out the intermediary and selling directly to the customer.

II. Literature Survey

Machine learning (ML) and its various applications in corporate technology, such as enhancing ERP system functionality and making data-driven decisions. The field of Artificial Intelligence (AI) known as ML enables computers to autonomously acquire new abilities and information from their own data sets [5]. Integrating ML algorithms with an ERP system allows for the rapid and accurate evaluation of large data sets. In [6], the authors provide a model for sales forecasting that makes use of the XGBoost algorithm in conjunction with careful feature engineering processing. The method outperforms the competing ML algorithms according to the experimental results. The RMSSE measure used in this study is 0.142 times lower than that of the Linear Regression (LR) method, and 0.114 times lower than that of the Ridge algorithm. In [7], Neural Network (NN) selling forecast algorithm is proposed. Similarly, the proposed systems use the Kaggle platform to assess NN models. Our NN model outperforms other ML models in all of our testing. Relative to the LR approach (2.93 times lower) and the Support Vector Machine (SVM) (2.55 times lower), our RMSE is significantly lower. In essence, [8] reviewed data and sales predictions. Later in the study, sales forecasting indicators and methods are detailed. The proposed systems describe outcomes using effective prediction and forecasting methodologies' reliability and accuracy. Results show Gradient Boost (GB) Algorithm is the best model for sales prediction. Sales forecasting analytics using ML are in [9]. Most research on building regression ensembles from separate models has focused on stacking. The findings suggest stacking may improve sales time series forecasting predictive models. In [10], ML methods are used to predict Rossman drugstore chain sales. This category also includes hybrid models like AutoRegressive Integrated Moving Average-Adaptive Recurrent Neural Network (ARIMA-ARNN), ARIMA-XGBoost, STL Decomposition, and ARIMA-SVM. The hybrid ARIMA-ARNN and ARIMA-SVM models outperformed their standalone counterparts. Decomposition and hybrid models were used [11]. This was followed by composite model STL development. Naive, ARIMA, and XGBoost predicted trend, residual, and seasonal components. [12] proved that SVM had better prediction accuracy than other conventional financial risk in ERP early warning systems by comparing and evaluating the two models. To show that machine learning can effectively predict financial crises in ecommerce, [13] used a Genetic Algorithm (GA) and a SVM to create an early warning model. The aforementioned ML algorithms are not capable of making any predictions using time series financial data, particularly with complicated data samples that have a long projection horizon. An Long Short-term Memory (LSTM) model is a kind of Deep learning (DL) that makes advantage of the long a high degree of predictability that is proportional to the separation between the two sets of data [14], which takes advantage of its built-in laws and works better with time series data. Natural language processing, finance, energy, and as well as traffic forecasting. By studying various ML methods, [15]found that LSTM, a DL algorithm model, outperformed the ARIMA model, an old-fashioned approach for ML. Researchers [16] combine LSTM with empirical modal decomposition (EMD) to input various sequences of all attributes, and the model did a fantastic job in trials that involved prediction. Using the Convolutional Neural Network (CNN) attention mechanism, [17] constructed a hybrid model. Using a Long Short-term Memory Bidirectional Neural Network (LSTMBiNN) to solve the problem of DOM prediction. Predictions reached 87% accuracy by the end. ERP software that incorporates AI may learn from its own data, leading to quicker and more accurate predictions and patterns. The classification method is employed to discover the company's brightest minds, say [18]. This prediction is based on an analysis of each worker's classification model performance. Researchers in many scientific fields used data mining classification techniques to generate new rules and make predictions. The scientists then propose an approach to performance prediction that they name EXGBRF, an acronym for Ensemble XG Boost RF Hybrid. The final result showcases absolute sensitivity [19]. Businesses can benefit from ERP systems because they allow for a more consolidated and dynamic view of key business processes, which in turn helps to streamline operations. ERP software ought to have the capability to integrate industry standards into the administration of several departments' work. These days, an integral part of any successful business is an ERP system [20]. Ensemble (Adaboosting and Gradient Boosting), k-nearest neighbor (KNN), LR, Decision Tree (DT), RF, and SVM are among the models. It was found that Gradient Boosting was the most effective of the classification methods that were examined. According to the findings, the promotion process was unaffected by bias and was unaffected by the characteristics department or the recruitment channel.

III. Methodology

A new manner of doing business, online shopping, has emerged as a result of the revolutionary impact of internet technology on people's daily lives. Online tools have made it easier than ever for companies to keep tabs on inventory in real time, facilitate better information sharing, and safeguard sensitive customer and staff information. Having effective and low-cost means of transferring information, funds, logistics, and human resources is critical for companies in this area.

A. Preprocessing

Data preprocessing is the method by which the chosen data is cleansed of any anomalies or background noise. This suggests that data that has a lot of extra significance is being cleaned up, even when it isn't needed. For instance, since product reviews aren't relevant to sales forecast in this dataset, we may safely assume that they are the source of the noise

and eliminate them. In addition, we must deal with the issue of missing sales and price values in the dataset in a proper manner by either imputation to the mean or median or by replacing the missing value with the average[21]. It might now also take into consideration the data's temporal sequence and the modifications that are known to have occurred.

1) Data Transformation

The term data transformation refers to the process of changing the format of data so that it can be used appropriately. This procedure is also known as the ETL process, which stands for Extract, Transform, and Load. As the volume of data has grown exponentially, transformation has emerged as a critical requirement. As a result, users will be able to zero in on data that caters to business requirements for robust data transformation. This project will also combine only the most crucial facts from the dataset after a thorough transformation. With this, the researcher can better zero in on the relevant data and construct a prediction model that yields more accurate results. The researcher must examine each of the eight datasets, extract relevant values and information, and then change the data's format to one or more others. Removing the need to examine superfluous data will greatly simplify the research process during the modelling stage.

B. Feature Selection:

1) KPCA:

The questionnaire has forty-two items that impact the ERP implementation process, derived from relevant literature and pilot survey conversations with working professionals. Using Principal Component Analysis (PCA), useful data predictions can be discovered[22]. With this basic transformation can be used to diagonalize an estimate of the covariance matrix of the data.

$$Z = \frac{1}{y}\sum_{v=1}^{y} q_l q_l^T \tag{1}$$

An eigenvector's principal components are its orthogonal projections. To deal with non-linear input data, Kernel PCA maps the space into feature space F by

$$\emptyset: \mathbb{R}^D \to H, q \to Q \tag{2}$$

On the newly altered feature space, the conventional PCA procedures are then executed. At its core, PCA aims to determine the eigenvalues $\mu \geq 0$ and the eigenvectors $K \in H\backslash\{0\}$ from the covariance matrix that satisfies Eq. (3) in the following way:

$$\bar{Z} = \frac{1}{y}\sum_{v=1}^{y} \emptyset(q_l)\emptyset(q_l)^T \tag{3}$$

For example, a kernel matrix (K) might be described as

$$V(q_u, q_l) := \emptyset(q_u)^T \emptyset(q_l) \tag{4}$$

In most cases, the inner products in the feature space are performed via a kernel function $V(q_u, q_l)$, rather than explicitly stating the transformation $\emptyset$.

C. Supervised Learning Models

1) Logistic Regression

Logistics regression, a soloist in the binary classification symphony, plays a starring role by outlining the ways in which ERP and e-commerce is being optimized. Akin to a virtuoso, this algorithm predicts the likelihood that a transaction falls into a particular category. It converts the linear combination of information into a melodic probability score by applying a transformation through the use of a logistic function. Intuitive and quick to process data, but it risks missing the mark when it comes to non-linear patterns due to its reliance on linear assumptions.

2) Support Vector Machines (SVM)

Visualize SVM's strength in the binary ballet of optimizing enterprise ERP and e-commerce—constructing hyperplanes in high-dimensional spaces—dividing classes. SVM creates strong separation by dancing around complicated decision limits in search of the hyperplane that maximizes the margin[23]. An expert in high-dimensional spaces with kernel tricks for non-linear storytelling, however its performance might be a bit shaky when dealing with larger datasets due to computational needs and kernel choices.

3) Random Forests and Decision Trees

A tree-like ballet develops while enter the whimsical forest of DT, where data is partitioned. As a big ensemble, random woods waltz in to strengthen predictions and stave off overfitting. With its interpretability and adaptability, DT dance through the numerical and categorical domains, revealing the complex decision boundary dance. On the one hand, DT can overfit, and on the other, random forests provide complexity, which can make the dance seem like a masquerade

4) Gradient Boosting Machines (GBM)

With the help of a symphony of weak learners, GBM leads the charge in this sequential ensemble learning crescendo. The virtuoso of capturing complex patterns in a dance where outliers are elusive, GBM improves with each model that comes before it. Although it excels in predictive accuracy and non-linear correlations, it is important to be cautious when tweaking hyperparameters to avoid overfitting. Reduces the majesty it brings to the stage because to training time and computational resources.

IV. Results and Discussion

Keeping up with the latest trends is crucial in the ever-evolving world of technology and business operations. The influence of these tendencies on ERP optimization is revealed in this review, which assesses the growth of ML integration with ERP systems. Integrating ML technology into ERP environments has made great strides in the past few years. Better forecasts and data-driven decisions are made possible by ERP systems by leveraging ML algorithms, which are known for their capacity to extract complex patterns from large datasets.

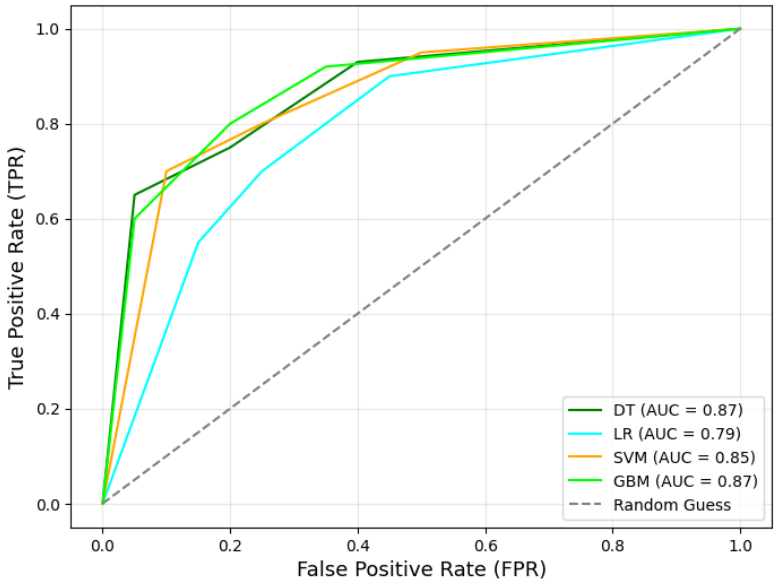

Fig. 1. ROC Comparison of Various Methods

Figure 1 shows the ROC curve, which shows how well different classification models—DT, LR, SVM, and GBM—performed on tasks related to optimizing Enterprise ERP and E-Commerce systems. For predictive tasks like demand forecasting, customer segmentation, and fraud detection, models with higher AUC values, like DT and GBM, are desirable. Through improved consumer retention and individualized recommendations, they elevate the e-commerce user experience. When it comes to driving decision-making and operational efficiency, the ROC analysis helps find the most successful models.

TABLE I. PERFORMANCE PREDICTION(%)

Metric	**Accuracy**	**Precision**	**Sensitivity**
LR	89.04	88.35	94.40
GBM	94.56	92.34	96.79
SVM	92.58	90.58	95.34
Decision Tree	90.63	89.36	93.87

Using the metrics of Accuracy, Precision, and Sensitivity, the table I compares the performance of four different models. For predictive tasks, Gradient Boosting Machine (GBM) is the best option due to its exceptional performance in terms of accuracy, precision, and sensitivity. Although it has the lowest metrics compared to the other models, Logistic Regression is still performing competitively, alongside SVM and Decision Tree. By comparing the two can see that GBM is the best model for sensitive and precision-sensitive applications like demand forecasting and fraud detection in enterprise resource planning and online store systems.

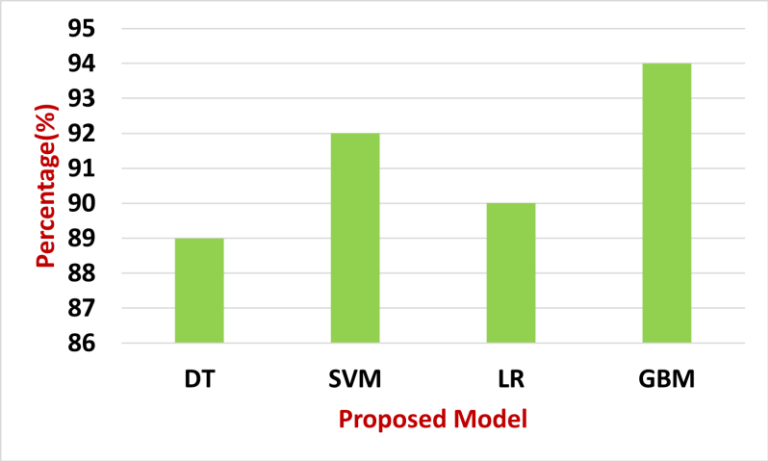

Fig. 2. Accuracy Comparison of Proposed Models

Based on their performance measures, DT, SVM, LR, and Gradient Boosting Machine (GBM) are compared in Fig. 2, a bar chart. By a wide margin, the other methods are surpassed by the proposed model, GBM, which achieves the highest percentage.

TABLE II. COMPARISON OF TRAINING AND TEST DATA (%)

Metric	Training Data	Testing Data
LR	89.04	87.35
GBM	94.21	93.03
SVM	91.76	89.22
Decision Tree	89.52	87.36

Models such as LR, DT, SVM, and GBM have their training and test datasets compared in Table 2.

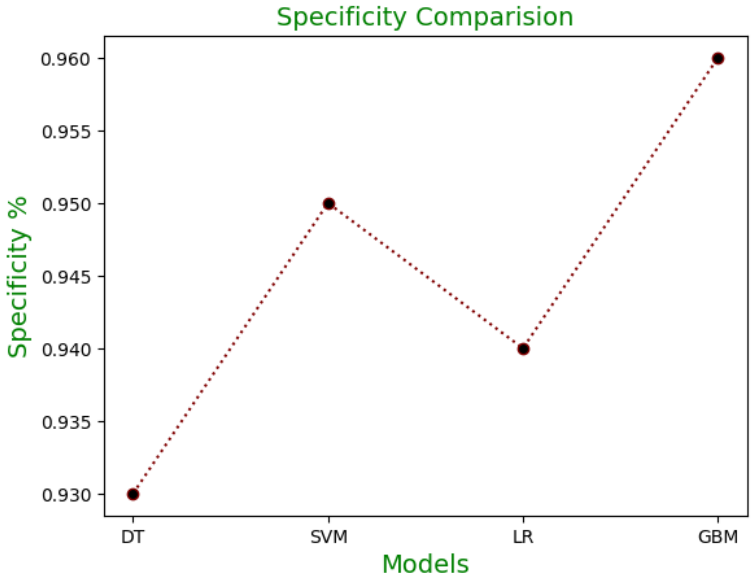

Fig. 3. Accuracy Comparison of Proposed Models

As shown in Figure 3, specificity is key when it comes to optimizing enterprise resource planning (ERP) and online store systems. This is because it guarantees the correct detection of non-critical events, including differentiating between legitimate and fraudulent transactions or non-essential changes to inventory. Given GBM's impressive performance, it could be a good fit for predictive jobs such as detecting fraud or segmenting customers. Results from models with modest computing efficiency and accuracy, such as SVM and LR, are indicative of possible trade-offs. By minimizing false alarms and optimizing ERP system procedures, high specificity has a direct effect on customer happiness in e-commerce.

V. Conclusion and Future Directions

In recent years, e-commerce platforms have become more accessible, allowing small and medium-sized businesses (SMEs) to offer their goods and services online. Creating an online storefront is simply one small part of digitalizing a business. Achieving digital transformation, as the saying goes, necessitates a complete overhaul of an organization's internal operations to accommodate data-driven decision-making, digital marketing tactics, and digital channel client involvement. Countless companies in the industrial sector have used ERP systems. Despite the abundance of success stories surrounding ERP deployment, many of these projects fall short of the organization's objectives, despite the fact that ERP has the ability to handle data fragmentation.

REFERENCES

[1] M. Wu, "Optimization of E-Commerce Supply Chain Management Process Based on Internet of Things Technology," *Complexity*, vol. 2021, 2021, doi: 10.1155/2021/5569386.

[2] P. K. Das, M. K. Majumder, B. K. Kaushik, and S. K. Manhas, "Integrating ERP with E-Commerce: A New Dimension Toward Growth for Micro, Small and Medium-Scale Enterprises in India," *Adv. Intell. Syst. Comput.*, vol. 236, no. SocProS, pp. 1117–1126, 2014, doi: 10.1007/978-81-322-1602-5.

[3] V. Kujala and R. Halonen, "Business growth using open source e-commerce and ERP in small business," *Adv. Intell. Syst. Comput.*, vol. 940, pp. 147–158, 2020, doi: 10.1007/978-3-030-16657-1_14.

[4] A. Kaya and Ö. Aydin, "E-commerce in Turkey and sap integrated e-commerce system," *Int. J. Ebus. eGovernment Stud.*, vol. 11, no. 2, pp. 207–225, 2019, doi: 10.34111/ijebeg.20191128.

[5] I. A. K. Shaikh, R. P. Pujar, S. P. Kishore, S. Ragamayi, P. V. Krishna, and A. B. Nadaf, "A Novel Approach for E-Commerce System for Sale Prediction with Denoised Auto Encoder and SVM based Approach," in *2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, IEEE, Jun. 2023, pp. 1684–1689. doi: 10.1109/ICSCSS57650.2023.10169595.

[6] R. Rajagopal, A. R. Arunarani, A. Arivarasi, A. Ingle, T. Ravichandran, and R. V. Prakash, "Enhanced Honey Badger Algorithm for Resource Allocation and Task Scheduling in Cloud Environment," *Proc. 4th Int. Conf. Smart Electron. Commun. ICOSEC 2023*, pp. 1375–1380, 2023, doi: 10.1109/ICOSEC58147.2023.10275908.

[7] S. Yadav, S. Mann, K. Renjith Krishnan, D. C. Dobhal, M. V. Unni, and S. Manoj, "The Performance Impact of Human Resources Recruitment System for Business Process Management Using K-Means and SVM," *Int. Conf. Self Sustain. Artif. Intell. Syst. ICSSAS 2023 - Proc.*, no. Icssas, pp. 701–706, 2023, doi: 10.1109/ICSSAS57918.2023.10331730.

[8] S. Yadav, R. Singh, E. Manigandan, M. V. Unni, S. Bhuvaneswari, and N. Girdharwal, "Research on Factors Affecting Consumer Purchasing Behavior on E-commerce Website During COVID-19 Pandemic based on RBF-SVM Network," *2nd Int. Conf. Autom. Comput. Renew. Syst. ICACRS 2023 - Proc.*, pp. 371–376, 2023, doi: 10.1109/ICACRS58579.2023.10404765.

[9] R. Umanesan, M. Shunmugasundaram, P. Rani, K. Balasubramanian, K. Rakesh, and P. S. Chandel, "Application of BiLSTM-CRF Approach and its Application for Better Decisions in Human Resource Management Processes," *Int. Conf. Sustain. Commun. Networks Appl. ICSCNA 2023 - Proc.*, no. Icscna, pp. 877–882, 2023, doi: 10.1109/ICSCNA58489.2023.10370629.

[10] K. M. Nayak, K. Renjith Krishnan, S. D. Khetree, M. V. Unni, A. Vignesh, and S. Ramya, "Fostering Synergy: Exploring the Intersection of Operations and Human Resources Management through an RF-SVM Approach," *2nd Int. Conf. Intell. Data Commun. Technol. Internet Things, IDCIoT 2024*, pp. 947–953, 2024, doi: 10.1109/IDCIoT59759.2024.10467998.

[11] K. Kumar *et al.*, "Forecasting E-Commerce Sales Adoption Based on DE-ELM-RGSO Approach," *2nd Int. Conf. Intell. Data Commun. Technol. Internet Things, IDCIoT 2024*, pp. 954–959, 2024, doi: 10.1109/IDCIoT59759.2024.10467941.

[12] N. Kogila, R. Rajkumar, S. Rajesh, and S. Vennila, "A Novel Approach of Ecommerce for Sales Prediction Using Hybrid ABC and AdaBoost Approach," in *1st International Conference on Electronics, Computing, Communication and Control Technology, ICECCC 2024*, IEEE, 2024, pp. 1–6. doi: 10.1109/ICECCC61767.2024.10593970.

[13] S. Vii, G. D. Rede, P. Ramesh, R. Kumar A, A. Bharathi, and M. C. J. Anand, "Optimizing E-Commerce Fraud Detection with BiGRU and Capsule Network Architectures," in *2nd IEEE International Conference on Data*

Science and Network Security, ICDSNS 2024, 2024, pp. 1–6. doi: 10.1109/ICDSNS62112.2024.10691229.

[14] S. Katuu, "Enterprise Resource Planning: Past, Present, and Future," *New Rev. Inf. Netw.*, vol. 25, no. 1, pp. 37–46, 2020, doi: 10.1080/13614576.2020.1742770.

[15] M. N. SALUR and W. K. KATTAR, "the Impact of Enterprise Resource Planning (Erp) on the Audit in the Context of Emerging Technologies," *Ekon. Maliye İşletme Derg.*, vol. 4, no. 2, pp. 115–123, 2021, doi: 10.46737/emid.1032735.

[16] O. V. Demyanova, E. V. Andreeva, D. R. Sibgatullina, A. M. Kireeva-Karimova, A. Y. Gafurova, and C. S. Zakirova, "Evaluation of effectiveness of information systems implementation in organization (by example of ERP-systems)," *J. Phys. Conf. Ser.*, vol. 1015, no. 4, 2018, doi: 10.1088/1742-6596/1015/4/042009.

[17] B. N. Prashanth and R. Venkataram, "Development of Modular Integration Framework between PLM and ERP Systems," *Mater. Today Proc.*, vol. 4, no. 2, pp. 2269–2278, 2017, doi: 10.1016/j.matpr.2017.02.075.

[18] S. Xu, J. Chen, M. Wu, and C. Zhao, "E-Commerce supply chain process optimization based on whole-process sharing of internet of things identification technology," *C. - Comput. Model. Eng. Sci.*, vol. 126, no. 2, pp. 843–854, 2021, doi: 10.32604/cmes.2021.014265.

[19] S. Li, "Structure Optimization of e-Commerce Platform Based on Artificial Intelligence and Blockchain Technology," *Wirel. Commun. Mob. Comput.*, vol. 2020, no. 3, pp. 1–8, Nov. 2020, doi: 10.1155/2020/8825825.

[20] S. Teerasoponpong and A. Sopadang, "Decision support system for adaptive sourcing and inventory management in small- and medium-sized enterprises," *Robot. Comput. Integr. Manuf.*, vol. 73, no. February 2021, p. 102226, 2022, doi: 10.1016/j.rcim.2021.102226.

[21] K. Singh, P. M. Booma, and U. Eaganathan, "E-Commerce System for Sale Prediction Using Machine Learning Technique," *J. Phys. Conf. Ser.*, vol. 1712, no. 1, 2020, doi: 10.1088/1742-6596/1712/1/012042.

[22] I. Ghosh and S. Biswas, "A novel framework of ERP implementation in Indian SMEs: Kernel principal component analysis and intuitionistic Fuzzy TOPSIS driven approach," *Accounting*, vol. 3, no. 2, pp. 107–118, 2017, doi: 10.5267/j.ac.2016.7.004.

[23] P. Kamuangu, "A Review on Financial Fraud Detection using AI and Machine Learning," *J. Econ. Financ. Account. Stud.*, vol. 6, no. 1, pp. 67–77, 2024, doi: 10.32996/jefas.2024.6.1.7.