Scientific Research Publishing

Journal of Software Engineering and Applications, 2024, 17, 203-213
https://www.scirp.org/journal/jsea
ISSN Online: 1945-3124
ISSN Print: 1945-3116

An Improved Enterprise Resource Planning System Using Machine Learning Techniques

Ahmed Youssri Zakaria, Elsayed Abdelbadea, Atef Raslan, Tarek Ali, Mervat Gheith, Al-Sayed Khater, Essam A. Amin

Department of Software Engineering, Faculty of Graduate Studies for Statistical Research (FGSSR), Cairo University, Giza, Egypt
Email: 12422022480767@pg.cu.edu.eg, e_abdelbadea@yahoo.com, Dr.Atef.Raslan@gmail.com, Tarekmmmmt@pg.cu.edu.eg, Mervat_gheith@yahoo.com, dr.elsied.khater@gmail.com, essam.amin@cu.edu.eg

How to cite this paper: Zakaria, A.Y., Abdelbadea, E., Raslan, A., Ali, T., Gheith, M., Khater, A.-S. and Amin, E.A. (2024) An Improved Enterprise Resource Planning System Using Machine Learning Techniques. *Journal of Software Engineering and Applications*, **17**, 203-213.
https://doi.org/10.4236/jsea.2024.175011

Received: April 18, 2024
Accepted: May 17, 2024
Published: May 20, 2024

Abstract

Traditional Enterprise Resource Planning (ERP) systems with relational databases take weeks to deliver predictable insights instantly. The most accurate information is provided to companies to make the best decisions through advanced analytics that examine the past and the future and capture information about the present. Integrating machine learning (ML) into financial ERP systems offers several benefits, including increased accuracy, efficiency, and cost savings. Also, ERP systems are crucial in overseeing different aspects of Human Capital Management (HCM) in organizations. The performance of the staff draws the interest of the management. In particular, to guarantee that the proper employees are assigned to the convenient task at the suitable moment, train and qualify them, and build evaluation systems to follow up their performance and an attempt to maintain the potential talents of workers. Also, predicting employee salaries correctly is necessary for the efficient distribution of resources, retaining talent, and ensuring the success of the organization as a whole. Conventional ERP system salary forecasting methods typically use static reports that only show the system's current state, without analyzing employee data or providing recommendations. We designed and enforced a prototype to define to apply ML algorithms on Oracle EBS data to enhance employee evaluation using real-time data directly from the ERP system. Based on measurements of accuracy, the Random Forest algorithm enhanced the performance of this system. This model offers an accuracy of 90% on the balanced dataset.

Keywords

ERP, HCM, Machine Learning, Employee Performance, Pythonista, Pythoneer

1. Introduction

Since the 1960s, when the first Enterprise Resource Planning system was developed, the emphasis has mostly been on streamlining and increasing efficiency of corporate processes through consolidation. These systems' basic architecture was predominantly data-driven, with a preference for using quantitative metrics to inform organizational choices [1].

Numerous studies have looked into how to incorporate machine learning into ERP systems with the goal of improving the efficiency of decision-making [2]. For instance, proposed an intelligent ERP system that uses machine learning for product quality prediction and supply chain optimization, while proposed a machine learning-based recommendation system for decision support in ERP systems [3].

Businesses today face a variety of difficulties in managing their supply chains, including flawed production planning, inaccurate demand forecasting, and ineffective inventory management. For production planning, proposed a particle swarm optimization algorithm that significantly decreased production costs. For inventory management, proposed a reinforcement learning strategy that produced sizable cost savings [4].

Data governance, skill sets, and model transparency are challenges of Machine Several studies have looked at how machine learning can be integrated with finance ERP. For instance, a framework for integrating machine learning with ERP was put forth by to improve the prediction of financial events like bankruptcies. To predict bankruptcies, they created models using machine learning algorithms and financial ratios obtained from ERP systems. Their findings demonstrated that the use of machine learning algorithms increased the predictability of financial events [5].

In recent times, the paradigm has shifted from a merely quantitative approach to a more qualitative approach, particularly in the domain of human resource management and employee evaluation [1]. As of right now, traditional methods of performance reviews are unable to provide timely feedback, track performance in real-time, or improve employee opportunities [6]. Furthermore, an employee's chances of being promoted may be harmed by a supervisor's prejudiced or false report. The bulk of organizations' existing promotion procedure should be viewed as misleading because it is reliant on supervisors' opinions [7].

Organizations must compare employee performance to expectations and put preventive and corrective measures in place to stop a decline in employee performance. It also gives the company the ability to inspire employees even more to enhance their productivity, working environment, and work skills. It also points out sections that require updating or changing. Employee attrition is the outcome of poor performance, and this will undoubtedly negatively impact the productivity and culture of the company [8].

Since it has a significant impact on employees' careers, performance, attitudes, loyalty, responsibilities, and general productivity, promotions are crucial in the

majority of organizations. In addition to other irreplaceable details, a corporation loses long-term employees' experiences and client relationships. We can ascertain whether a specific employee has shown a pattern of improvement based on previous performance evaluations. Inaccurate information could lead to the reward of unfit workers rather than diligent ones. The challenging challenge of deciding which personnel should be promoted faces many organizations [7].

Regardless of the scheduled assessment times (annual or quarterly), we explored in this study the significance of integrating machine learning (ML) with enterprise resource planning (ERP) systems and leveraging Python capabilities to use ML algorithms on ERP real-time data to anticipate employee evaluation at any time.

As businesses aim to use technology to discover new efficiencies and insights, using Machine Learning techniques in ERP systems offers a promising way forward. Incorporating ML methods in ERP systems provides a strategic opportunity to improve the precision and effectiveness of salary forecasts, helping organizations enhance compensation strategies and stay competitive in the talent market. Machine learning algorithms can detect employees who may leave their jobs because they are dissatisfied with their salary, allowing for preemptive actions to keep them [9].

By using both machine learning (ML) and Enterprise Resource Planning (ERP) together, organizations can predict employee salaries more effectively. This approach involves analyzing various data points, including education, experience, job position, and performance metrics, to estimate an employee's salary accurately. This approach to data utilization ensures that salaries are fairly and impartially determined based on appropriate criteria. Machine learning algorithms can analyze historical salary data and market trends to provide insights into typical salary ranges for different roles and locations. This information can help HR professionals make improved decisions when developing compensation strategies and adjusting salary structures to remain competitive in the market [10].

Machine learning algorithms help identify and minimize pay disparities and biases by analyzing salary data and identifying discrepancies related to attributes like gender, race, or other demographics. This can help organizations ensure fair and impartial compensation practices. Accurately forecasting employee salaries can help identify workers who may quit due to being unhappy with their pay, allowing companies to proactively intervene to retain important staff members. Overall, employing machine learning to predict employee wages in human capital management may lead to better, impartial, and data-driven payment methods that benefit both employees and the organization as a whole [11].

Multitudinous ERP providers have started to add AI capabilities to their systems in order to add AI-powered features. These enterprises demonstrate that incorporating AI into ERP systems is possible and act as early pointers of AI readiness. Still, to use the AI-driven pall services offered by these merchandisers,

the stoner must subscribe to SaaS. also, to customize any of the products like SAP S/4HANA, Oracle Fusion, Microsoft Dynamics 365, and In for Cloud-Suite—the client must subscribe to PaaS. Likewise, seller restrictions and limits apply to AI-driven functionality.

Without the need for additional infrastructure or programming expertise, the study enabled Pythonista (application users) to run machine learning algorithms in their environment.

2. Background

1) Employee performance evaluation is critical in any organization, as it has a significant effect on employee career performance and company productivity. The HR department companions the employees to achieve their career thing within organizations. Consequently, an organization builds an educated pool by retaining effective employees, who, in turn, become competent leaders in the future [7].

2) Enterprise Resource Planning (ERP) systems provide an integrated and continuously streamlined view of core business processes, allowing organizations to optimize their operations. The ERP system's function is to apply best practices to the management of numerous departmental responsibilities inside a business. ERP systems are currently one of the most crucial elements of any successful company and are growing more and more essential [1]. Popular ERP systems include Oracle EBS, SAP, and Microsoft Dynamics.

3) Artificial Intelligence (AI) The form of intelligence that is represented by a computer or software is called Artificial Intelligence (AI). It might include everything from simple automated processes to extremely complex systems that can forecast results through data analysis. The preferred method for developing software or computers that can grasp, learn from gathered data, and make judgments based on patterns that the computer can recognize is artificial intelligence (AI). These patterns might be a little too dramatic for humans to notice or quickly and accurately calculate on their own [1].

4) The machine learning (ML) division of AI helps to automate the process of building logical models. It is based on the idea that computers are capable of learning from data, identifying patterns in it, and forming conclusions with minimal assistance from humans [7].

5) A Random Forest is a model that averages the predictions made by numerous decision trees. These trees are produced by varying several factors, including the data used to train the tree and other tree attributes [7].

6) A Pythonista is a somebody who uses the Python programming language to do out tasks. Those that are continuously developing new ideas and using Python to create new things are known as Pythonista [12].

7) Pythoneer is the most popular term for an experienced Python coder. Because of the expertise of Python experts, the term "Pythoneer" can be used more broadly [12].

3. Related Work

R, S Dhivya and P, Sujatha [6] in 2023 mentioned, that the classification approach is used to identify talented employees within the organization. The classification model is used to determine whether a given hand performs well or poorly in order to make this forecast. Scientists in a variety of scientific domains generated and predicted rules using data mining classification algorithms. Next, they suggest using the Ensemble XG Boost Random Forest Hybrid Algorithm (EXGBRF) to predict performance. The outcome demonstrates complete delicacy [6].

Alqahtani, Fatma Ayed, and, Abdulaziz Almaleh in 2022 [7] mentioned that the prediction models listed below were created. Models such as KNN, Ensemble (Adaboosting and Gradient Boosting), Random Forest, Decision Tree, KNN, Support Vector Machine, and Logistic Regression. The Gradient Boosting algorithm fared better than the other classification algorithms, according to the results. The findings show that there was no prejudice and that promotion was not significantly influenced by the department of characteristics or the recruitment channel. Out of all the features, the ratings from the previous year were the most significant. More information would enable more refined solutions. The goal of this study was to use machine learning to identify which exceptional workers would be eligible for promotions. The primary objective was to build prediction models with classification algorithms in order to identify the most important variables affecting employee advancement and to forecast if a worker is eligible for one [7].

Goundar, Sam; Nayyar, Anand; Maharaj, Moniker; Ratnam, Karunesh; Prasad, Shalvin in 2021 [1] mentioned, the subsequent inquiries for research: What issues do the ERP systems have? How might AI address these issues? The study's conclusion is that knowledge management and organizational decision-making processes can be enhanced by combining AI and decision support systems with IT systems. By evaluating greater datasets than previously allowed, AI in ERP systems can enhance modular functionalities. AI is incorporated into ERP systems to enable the software to learn from previous data collected by the system and train itself, resulting in the provision of patterns and forecasts more quickly and accurately [1].

Nasr, Mona; Shaaban, Essam; Samir, Ahmed in 2019 [13] mentioned that management takes an interest in employees' work to guarantee that the appropriate person is put in the right position at the correct time. Since then, data mining (DM) has played an increasingly important part in the process of gaining knowledge from vast volumes of data. An actual dataset was supplied by the Ministry of Egyptian Civil Aviation (MOCA) in order to forecast the classification model used in this paper, which was built using DM techniques. Among the techniques are Decision Trees (DT), Nave Bayes, and Support Vector Machines (SVM). With the best prediction accuracy across all three studies, the SVM approach emerged as the most successful classifier for creating the pre-

dictive model [13].

4. Methodology

As represented in Figure 1, the following are the main stages of the suggested methodology for compensation or performance predictions for employees. During the first stage, known as the "Build" step, the Pythoneer constructs the model and examines a collection of variables from the training dataset. It is

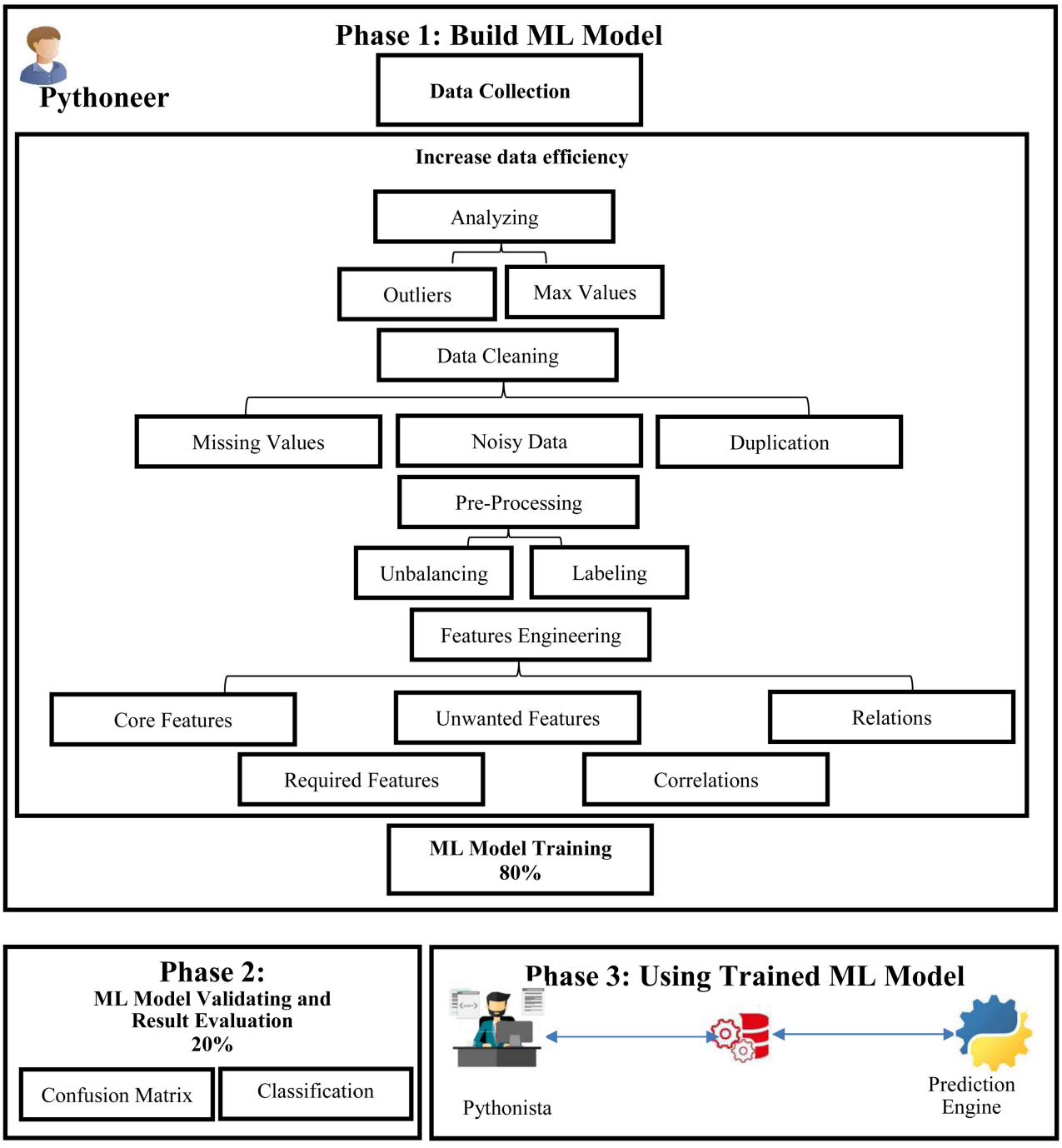

Figure 1. The overall proposed model.

presumed that every variable is related to and pertains to a designated class. By testing the model with a different dataset, the second phase is in charge of determining the model's or classifier's correctness, or validating the model. The final phase is when Pythonista can apply the model or classifier to newly discovered data to make predictions about particular unknown label classes if the classifier accuracy is deemed acceptable. There are three main stages to the first phase. The gathering of data from ERP is the initial step. Enhancing data efficiency represents the second phase. In order to prepare the data for future analysis, this step involves identifying and handling missing values, eliminating noisy, unnecessary, or duplicate distribution data, and handling imbalanced distribution data. Pythonista selects features at this stage as well in order to reduce unnecessary data and eliminate undesirable qualities in order to prevent overfitting. The final step involves using Pythoneer to feed our suggested Random Forest model with the train and test data in order to forecast employee performance.

4.1. STAGE 1—Data Collection

Finding a model that can predict employee performance using real-time data straight from ERP is the aim of this research. The "oracledb" library is utilized in our prototype to extract data from Oracle EBS "HCM Performance Management Application" and subsequently transform it into a DataFrame format. Table 1 illustrates the 13 features present in the data. The columns in the test and train data are the same, with the exception of is_promoted, which is missing from the test data.

Table 1. Employee dataset features for employee.

Attributes	Data Type	Description
employee_id	int64	Unique ID for employee
Department	String	Department of employee
Region	String	Region of employment (unordered)
Education	String	Education Level
Gender	String	Gender of Employee
recruitment_channel	String	Channel of recruitment for employee
no_of_trainings	int64	no other training completed in the previous year on soft skills, technical skills, etc
Age	int64	Age of Employee
previous_ year_ rating	float64	Employee Rating for the previous year
length_ of_ service	int64	Length of service in years
awards_ won?	int64 1 Yes, 0 No	if awards were won during the previous year then 1 else 0
avg_training_score	int64	Average score in current training evaluations
is_promoted	int64 1 Yes, 0 No	(Output) Eligible for promotion

4.2. STAGE 2—Increase Data Efficiency

Pre-processing at this point eliminates noisy data and improves the efficacy of the data. Subsequently, the data are deconstructed and forwarded to subsequent phases. Avoiding unnecessary and gratuitous variables is essential to ensure the model operates effectively. Another method of data reduction is thought to be data conception. An outlier can only exist in the length_of_service and avg_training score columns, just like in Table 1. Given that other variables can have many values.

Is_promoted is the desired column. How uneven it is shown in Figure 2. Stated otherwise, the ratio of 0 to 1 values is higher. This could lead to the model becoming biased. To address this, we employed Random Over Sampler, which has the advantage of increasing the quantity of samples the model needs to train on.

Figure 3 and Figure 4 present samples of data before and after Stage 2, while we:

- Fill the missing values in education and previous_year_rating;
- Create a new features (sum_metric and total_score);
- Remove unnecessary features (employee_id);
- Apply Label Encoder for Normalization on department, gender, and recruitment_channel.

4.3. STAGE 3—ML Model

For this investigation, a Random Forest model was utilized. With this method,

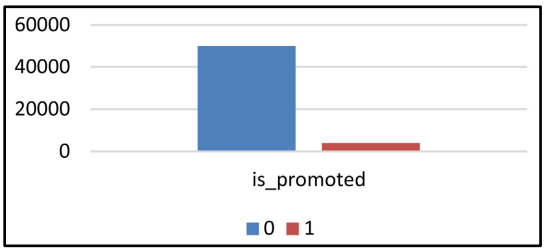

Figure 2. Distribution of target variable.

employee_id	department	region	education	gender	recruitment_channel	no_of_trainings	age	previous_year_rating	length_of_service	awards_won?	avg_training_score	is_promoted
65438	Sales & Marketing	region_7	Master's & above	f	sourcing	1	35	5	8	0	49	0
65141	Operations	region_22	Bachelor's	m	other	1	30	5	4	0	60	0
7513	Sales & Marketing	region_19	Bachelor's	m	sourcing	1	34	3	7	0	50	0
2542	Sales & Marketing	region_23	Bachelor's	m	other	2	39	1	10	0	50	0
48945	Technology	region_26	Bachelor's	m	other	1	45	3	2	0	73	0
...	...	...	...	...	...	...	...	...	...	...	...	...
74615	R&D	region_31	Bachelor's	m	sourcing	1	30	NaN	1	0	88	0
11685	Operations	region_15	Bachelor's	m	sourcing	1	31	NaN	1	0	56	1
10546	Finance	region_6	Bachelor's	m	other	1	28	NaN	1	0	61	0
37919	Finance	region_2	Bachelor's	m	other	1	23	NaN	1	0	61	0
12431	Technology	region_26	Bachelor's	f	sourcing	1	31	NaN	1	0	78	0

Figure 3. Sample of dataset before Stage 2.

department	education	gender	recruitment_	no_of_trainir	age	previous_ye	length_of_se	awards_won	avg_training	is_promoted	sum_metric	total_score
7	3	0	2	1	35	5	8	0	49	0	5	49
4	2	1	0	1	30	5	4	0	60	0	5	60
7	2	1	2	1	34	3	7	0	50	0	3	50
7	2	1	0	2	39	1	10	0	50	0	1	100
8	2	1	0	1	45	3	2	0	73	0	3	73

Figure 4. Sample of dataset after Stage 2.

```
print("Training Accuracy :", model.score(x_train, y_train))
print("Testing Accuracy:", model.score(x_valid, y_valid))

Training Accuracy : 0.9932688472277623
Testing Accuracy: 0.9756182688472278
```

Figure 5. Model scores.

	precision	recall	f1-score	support
0	1.00	0.95	0.97	9981
1	0.95	1.00	0.98	10075
accuracy			0.98	20056
macro avg	0.98	0.98	0.98	20056
weighted avg	0.98	0.98	0.98	20056

Figure 6. Classification matrix.

data is seeded into our suggested prediction model after training and testing are split into 80:20 ratios. Among the models at our disposal is this one. We presented this model solely as a proof-of-concept for the investigation.

4.4. Evaluation

Similar to Figure 5, 99% of the training data's items were anticipated correctly. Despite the high result, there is no reason to believe that overfitting occurred because the accuracy score for the test data is similarly high. In other words, 98% of the test data's entries were accurately anticipated.

Similar to Figure 6, the model accurately predicts 98% of the cases in the dataset based on its 0.98 accuracy, recall, macro average precision, and F1-score.

5. Results and Discussions

This study suggests that we may integrate AI capabilities into ERP systems on-premises, eliminating the requirement for cloud servers and vendor restrictions. Additionally, it can be customized by using Python's open committee support.

In order to maintain relevance and competitiveness in dynamic corporate contexts, machine learning (ML) algorithms in ERP systems evaluate large amounts of data, enhance decision-making, automate repetitive operations, and continuously learn from new data. This essay has demonstrated and examined the idea that an organization's real-time data performs best when predictions are accurate, and it has been shown to be done so in good faith.

6. Conclusion

The suggested approach for forecasting employee performance is broken down into three main stages, as shown in Figure 1. The Pythoneer constructs the model and examines a collection of training dataset variables in the first stage, which is referred to as the build step. Additionally, the Pythoneer in the second

phase is in charge of determining the model's or classifier's correctness (validating the model) by putting it to the test on a different dataset. Pythonista can apply the model or classifier to newly discovered data to create predictions by submitting a request straight from the ERP application with the employee's number or name, and this is regarded as the third phase, provided the classifier's precision is deemed appropriate.

Acknowledgements

The Pythonista Egypt Committee's support was essential to making this work possible. We sincerely appreciate the committee's provision of the necessary academic time.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Goundar, S., Nayyar, A., Maharaj, M., Ratnam, K. and Prasad, S. (2021) How Artificial Intelligence is Transforming the ERP Systems. *Enterprise Systems and Technological Convergence*: *Research and Practice*, South Pacific, Information Age Publishers, 1-15.

[2] Chopra, R., Sawant, L., Kodi, D. and Terkar, R. (2022) Utilization of ERP Systems in Manufacturing Industry for Productivity Improvement. *Sciencedirect*, **62**, 1238-1245. https://doi.org/10.1016/j.matpr.2022.04.529

[3] Chaoliang, H. and Qi, Z. (2021) Optimization of Supply Chain Efficiency Management Based on Machine Learning and Neural Network. *Neural Computing and Applications*, **33**, 1-10.

[4] Yan, A.Y., Suraksha, G.B., Tana, C.L.C. and Klaus, S.B. (2022) Integrating Machine Learning, Modularity and Supply Chain Integration for Branding 4.0. *Sciencedirect*, **104**, 136-149. https://doi.org/10.1016/j.indmarman.2022.04.013

[5] Iota, K.N. and Robert, P. (2021) Artificial Intelligence, Machine Learning and Big Data in Finance. OECD. https://www.oecd.org/finance/financial-markets/Artificial-intelligence-machine-learning-big-data-in-finance.pdf

[6] Dhivya, R.S. and Sujatha, P. (2023) Employee Performance Prediction for Workforce Planning using Ensemble Hybrid Model. In 2023 10*th International Conference on Computing for Sustainable Global Development* (*INDIACom*), New Delhi.

[7] Alqahtani, F.A. and Almaleh, A. (2022) Analysis and Prediction of Employee Promotions Using Machine Learning. 2022 5*th International Conference on Data Science and Information Technology* (*DSIT*), Shanghai, 22-24 July 2022. https://doi.org/10.1109/DSIT55514.2022.9943959

[8] Alsobaey, T.M. and Al-Alawi, A.I. (2023) The Use of Data Mining Techniques to Predict Employee Performance: A Literature Review. In 2023 International Conference On Cyber Management and Engineering (CyMaEn), Bangkok, 26-27 January 2023. https://doi.org/10.1109/CyMaEn57228.2023.10050937

[9] Kimberling, E. (2024) The Benefits of ERP for Human Resources Management.

https://www.thirdstage-consulting.com/erp-for-human-resources-management/

[10] Zambuk, F.U. and Abdullahi, A. (2023) Salary Prediction Model Using Principal Component Analysis and Deep Neural Network Algorithm. *International Journal of Innovative Science and Research Technology*, **8**, 10.

[11] Shaprapawad, S. and Borugadda, P. (2024) Education and Society Salary Prediction Using Machine Learning Regression Algorithms. *Education and Society*, **48**, 88-94.

[12] Ashwin, J. (2021) What Is a Pythonista and How to Become One. https://pythonistaplanet.com/pythonista/

[13] Nasr, M., Shaaban, E. and Samir, A. (2019) A Proposed Model for Predicting Employees' Performance Using Data Mining Techniques: Egyptian Case Study. *International Journal of Computer Science and Information Security* (*IJCSIS*), **17**, 1-10.