Assessing Data Analytics Capabilities in Retail Organizations: Insights into Mining, Predictive Analytics and Machine Learning

ROSARIO PARIONA-LUQUE[1, a], ALEX PACHECO[2, b], EDWIN VEGAS-GALLO[3, c], RUI ALEXANDRE CASTANHO[4,5,6,d], FABIAN LEMA[1,e], LIZ PACHECO-PUMALEQUE[7,f,*], MARCO AÑAÑOS-BEDRIÑANA[8,g], WILSON MARIN[9,h], EDWIN FELIX-POICON[10,i] ANA LOURES[4,j]

[1]Sustainable Tourism and Hospitality Management,
Universidad Nacional Autonoma de Huanta,
Jr. Manco Cápac 497, Huanta, Ayacucho,
PERU

[2]Professional School of Systems Engineering,
Universidad Nacional de Cañete,
Av. Benavides 1370, San Vicente de Cañete, Lima,
PERU

[3]Universidad Peruana de Ciencias e Informática,
Av. Talara 752 Jesús María, Lima,
PERU

[4]VALORIZA - Research Center for Endogenous Resource Valorization,
Instituto Politécnico de Portalegre (IPP),
PORTUGAL

[5]College of Business and Economics,
The University of Johannesburg, P.O. Box 524 Auckland Park,
Johannesburg 2006,
SOUTH AFRICA

Advanced Research Centre,
[6]European University of Lefke, Lefke, Northern Cyprus,
TR-10 Mersin,
TURKEY

[7]Carrera de Administración, Facultad de Ciencias Empresariales,
Universidad San Ignacio de Loyola,
Av. La Fontana 550, La Molina, Lima,
PERU

[8]Universidad Nacional Autónoma de Chota,
Jr. José Osores 418, Chota, Cajamarca,
PERU

[9]Universidad Cesar Vallejo,
Av. Victor Larco 1770, Trujillo, La Libertad,
PERU

[10]Universidad Nacional de Jaen,
Carretera Jaén - San Ignacio KM 24 - Sect. Yanuyacu, Jaén, Cajamarca,
PERU

[a]ORCiD: https://orcid.org/0000-0002-8468-7801
[b]ORCiD: https://orcid.org/0000-0001-9721-0730
[c]ORCiD: https://orcid.org/0000-0002-2566-0115
[d]ORCiD: https://orcid.org/0000-0003-1882-4801
[e]ORCiD: https://orcid.org/0000-0002-0168-8703
[f]ORCiD: https://orcid.org/0000-0002-4323-1293
[g]ORCiD: https://orcid.org/0000-0002-4737-4443
[h]ORCiD: https://orcid.org/0000-0001-6175-8112
[i]ORCiD: https://orcid.org/0000-0001-5536-2410
[j]ORCiD: https://orcid.org/0000-0002-2146-4205

Abstract: - Nowadays, implementing data analytics is necessary to improve the collection, evaluation, analysis, and organization of data that allow the discovery of patterns, correlations, and trends that improve knowledge management, development of strategies, and decision-making in the organization. Therefore, this study aims to provide an accurate and detailed assessment of the current state of data analytics in the retail sector, identifying specific areas of improvement to strengthen knowledge management in organizations. The research is applied with a quantitative approach and non-experimental design at a descriptive and propositional level. The survey technique was used, and as a data collection instrument, a questionnaire addressed to 351 employees of companies in the retail sector concerning the variable data analysis with the dimensions of data extraction, predictive analysis, and machine learning and the variable management of the knowledge with the dimensions knowledge creation and knowledge storage. The results show that 52.99% of collaborators indicate that the level of data extraction is terrible, 57.83% indicate that the level of predictive analysis is wrong, and 54.99% express that the level of machine learning is average, which contributes to the implementation of innovative resources and solutions that promote the inclusion of a high-tech approach to address information management problems and contribution to the development of knowledge in an institution.

Key-Words: - Data analytics; knowledge management; Data extraction; predictive analytics; machine learning; knowledge creation; storage of knowledge.

Received: October 16, 2023. Revised: May 11, 2024. Accepted: June 12, 2024. Published: July 5, 2024.

1 Introduction

Nowadays, data analytics is becoming a fundamental practice in extracting information as it allows the use of relevant and quality data to improve the decision-making process in organizations, [1]. Also, technology is a critical element in data analytics as it allows the creation of platforms that provide modern data analysis functions and patterns; however, they can generate a counterproductive effect if the platform provides inadequate or difficult-to-understand information [2], [3]. Similarly, the benefits of data analytics are revenue growth, cost savings, increased market share, customer satisfaction and corporate market value gains, and the discovery of hidden patterns, unknown correlations, trends, and customer preferences, [4]. Finally, data analytics promotes quality improvement, process optimization, and early detection of parameter deviation, which provides an economic advantage over other organizations in terms of productivity and efficiency gains, [5].

In the business sector, the enhancement of data analytics is not just a choice, but a necessity. It empowers organizations to bolster their knowledge management, producing quality reports that significantly enhance decision-making capabilities. As [6] suggests, knowledge management is a process of creating, organizing, evaluating, transmitting, and applying knowledge to solve practical problems. Similarly, [7], outlines that knowledge management involves planning, conducting, monitoring, and evaluating actions and decisions related to the acquisition, transmission, preservation, retrieval, creation, application, and dissemination of data, information, and knowledge. [8], further argues that knowledge management is essential in organizations to foster the creation of new knowledge, share information with all staff, and improve employee performance. The benefits of knowledge management are profound, strengthening the organization's flexibility by promoting the free dissemination of information, increasing institutional value, and improving competitiveness, [9].

Despite the undeniable benefits of data analytics, significant challenges remain in the practice globally. In China, problems such as low accuracy in analysis, incomplete information extraction, and poor implementation of traditional analytics methods have been identified, restricting the effectiveness of data analytics in formulating competitive strategies in organizations, [10]. Similarly, in Malaysia, there is evidence of inadequate student preparation in data analytics, highlighting the need to enhance specialized skills in data visualization, data storytelling, and the handling of data analytics tools, [11]. In addition, India recognizes the urgent need for some companies to implement technologies in data analytics to manage, analyze, store, and process massive volumes of data in shorter times, [12]. Finally, in Colombia, it is reported that many companies need help in decision-making and strategy formulation due to the limited capacity to process the overwhelming amount of data generated by the market, highlighting the imperative need to adopt new technologies, [13].

1.1 Knowledge Gap, Objectives, and Scope of the Research

Previous studies have found that implementing data analytics is essential to optimize data collection, evaluation, and organization processes, thus facilitating the identification of patterns, correlations, and trends crucial for effective knowledge management. In this regard, some case studies have shown that companies that effectively incorporate data analytics experience significant improvements in strategic decision-making and operational efficiency.

However, the literature needs more solid, correct, and up-to-date evidence on the specific perception and application of data analytics in the retail sector and its direct relationship with knowledge management. This paper aims to fill this gap by analyzing in detail the level of implementation of data analytics in retail companies, focusing on critical dimensions such as data mining, predictive analytics, and machine learning and their impact on knowledge management.

Contextually, this paper shows that more than half of the surveyed employees express dissatisfaction with the level of data mining, predictive analytics, and machine learning in their organizations. Therefore, there is a clear need to implement innovative resources and solutions that integrate a high-tech approach to effectively address information management challenges and contribute to developing knowledge in the business environment.

Therefore, this study aims to provide an accurate and detailed assessment of the current state of data analytics in the retail sector, identifying specific areas of improvement to strengthen knowledge management in organizations. This research contributes to the field by providing a deeper understanding of the retail sector's challenges and opportunities in implementing data analytics, highlighting the importance of innovative strategies for data-driven decision-making and organizational knowledge growth.

2 Literature Review

The comprehensive literature review explores the fundamental dimensions of data analytics and knowledge management. In this context, we focus on the importance of data mining, predictive analytics, and machine learning in data analytics.

First, the data mining dimension of the data analytics variable is conceptualized as a process responsible for extracting and compiling information from semi-structured and unstructured sources to streamline data analysis and reporting processes, [14]. Likewise, it is known that to improve data extraction, it is essential to work with specialized programs that include the construction of the sensor hardware, the development of the perception algorithm, and the scenario data, [15]. In that sense, a study conducted in Australia developed a data mining algorithm that allowed them to improve data processing through flexible, efficient, and accessible data mining, [16].

Secondly, [17], indicate that predictive analytics is an advanced study method responsible for deeply examining data, reports, and content to predict some market sector's risks, opportunities, or behaviors. Likewise, [18], mentions that nowadays, predictive analytics uses statistical modeling techniques and new technologies, such as big data and machine learning, in the elaboration of predictions to achieve continuous performance improvement and information flow management. Similarly, research on oil handling operations in Russia showed that predictive analytics helps detect environmental risks and operating personnel's physical condition, [19].

Machine learning, [20], points out that it is a branch of artificial intelligence that endows computers with the ability to learn and act according to the needs and information about the situation. In addition, [21], [22], states that machine learning works based on computer programs that allow the performance of non-explicitly programmed actions

based on the information available and the patterns found in the data. A study in China, which included machine learning for forest fire forecasting, showed that it is helpful for disaster prediction, considering each region's particular characteristics, [23].

On the other hand, the knowledge creation dimension of the knowledge management variable is defined as a collective process involving actors participating in exchanging and integrating different knowledge to realize innovative ideas, [24], [25]. Similarly, knowledge creation is essential for organizations to achieve continuous improvement and become more competitive in the market, as the new strategies and innovations generated allow them to satisfy customers in the face of market changes, [26], [27]. In that framework, research in Finland stated that knowledge creation requires the analysis of past and present data to develop new knowledge that better understands customer needs, [28].

According to [29], knowledge storage involves organizing and distributing knowledge in various databases, intranets, extranets, and information systems that enable organizations to have a knowledge map. Likewise, [30], argues that knowledge storage is fundamental for the consolidation of an organization's knowledge, as it allows new theories, patterns, ideas, and information to be stored, creating a collaborative network that enables the institution's workers to interact with the information in order to increase the level of productivity. In this sense, a study conducted in Thailand showed that organizations need to carry out knowledge storage to have the necessary information available to innovate, lead, and direct strategies, [31].

3 Methodology

3.1 Design

The planning of this research, which is applied and non-experimental, focused on a correlational-causal approach. This enabled a detailed exploration of the interrelationships between the various dimensions of data analytics and the critical aspects linked to knowledge management, [32]. This methodology was chosen due to its ability to identify patterns and connections inherent in the data without disturbing the participants' natural environment, thus preserving the authenticity of the work context in the construction industry.

3.2 Inclusion and Exclusion Criteria

The study's intentional sample comprised 351 collaborators, distributed between 189 men and 162 women. Thereby, to obtain this result, specific inclusion criteria were applied: (a) the age of the participants had to be between 25 and 50 years, (b) they had to give their consent to participate in the research, and (c) they were required to be permanent workers with at least ten months of work experience in the retail sector. The choice of this sector for research is due to its relevance and significant presence in the business environment, providing an ideal context to examine the implementation of data analytics in a dynamic and competitive business environment. Exclusion criteria, on the other hand, included (a) submission of incomplete questionnaires and (b) unwillingness to continue participating in the study, ensuring the quality and consistency of the data collected.

3.3 Procedure

The research was carried out from October to December 2023, during which the participants were recruited continuously using convenience sampling until they reached the desired sample size (351 workers), considering there was no incomplete questionnaire. The data collection technique was the survey, applying a structured questionnaire using the Google Forms tool to measure opinion about the data analytical variable according to its dimensions: data extraction, predictive analysis, and machine learning with a total of 15 questions and ten questions to measure opinion regarding the knowledge management variable and the dimensions: knowledge creation and knowledge storage, using the Likert scale according to the values good, average and bad. The methodology of this study emphasizes a detailed understanding of the various dimensions of data analytics, integrating them as central axes in the analysis. Integrating these three dimensions in our methodological approach allows a comprehensive evaluation of how data analytics contributes to improving knowledge management in companies in the retail sector. By focusing our analysis on these dimensions, we seek to offer a comprehensive perspective on the benefits and challenges of properly implementing data analytics, highlighting its potential to improve efficiency and effectiveness in knowledge management.

3.4 Analysis of Data

This study organized the collected data into a tabulation matrix and processed it using SPSS v25 and Excel statistical software. In this regard, the

dimensions of data analytics (data mining, predictive analytics, and machine learning) should be measured. Cronbach's Alpha reliability tests were conducted for the variables related to data analytics and knowledge management, resulting in a coefficient of 0.911. This value reflects a high reliability in the measurements, guaranteeing the internal consistency of the answers collected through the questionnaire. During the research process, descriptive statistics were applied to perform the frequency distribution of the dimensions related to data mining, predictive analytics, and machine learning, as well as the dimensions of knowledge creation and storage in organizations. This approach provided a clear understanding of the general trends in participants' perceptions and experiences regarding the implementation of data analytics.

3.5 Ethical Considerations

Ethical principles of research were followed, ensuring confidentiality and informed consent for all participants. Personal information and responses were handled confidentially and used exclusively for research purposes.

4 Findings

Figure 1 shows the results of the data mining dimension of the data analytics variable. Some 52.99% of contributors indicate that the level of data mining could be better, suggesting a severe deficiency in the organizations' ability to obtain relevant information from their systems. This poor data mining performance can significantly limit the effectiveness of subsequent analysis and informed decision-making. Only 26.21% consider the level reasonable, indicating that less than a third of employees perceive that their organizations adequately handle this crucial stage of the analytical process. In addition, 20.80% of employees say the level is fair, reflecting a general perception that there is much room for improvement in this area.

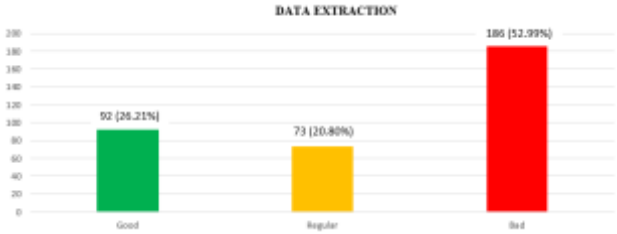

Fig. 1: Data extraction dimension level

Figure 2 shows the results of the predictive analytics dimension of the data analytics variable, with 57.83% of employees indicating that the level of predictive analytics is poor. Thus, this indicates that most employees perceive their organization's need to leverage predictive tools to anticipate future trends and behaviors. The fact that only 18.52% consider the level to be good underlines a worrying lack of confidence in current predictive analytics capabilities. 23.65% of respondents believe the level is fair, suggesting that while some organizations are on the right track, they still face challenges in achieving an optimal level of predictive analytics.

Fig. 2: Level of the predictive analytics dimension

Figure 3 shows the results of the machine learning dimension of the data analytics variable. 54.99% of respondents indicate that machine learning is fair, suggesting that, although organizations use this technology, its implementation and effectiveness are not consistently high. Furthermore, 23.93% of respondents consider the level poor, indicating that essential areas need significant improvement. Only 21.08% say the level is good, highlighting the need to strengthen machine learning capabilities so that organizations can fully benefit from its advantages.

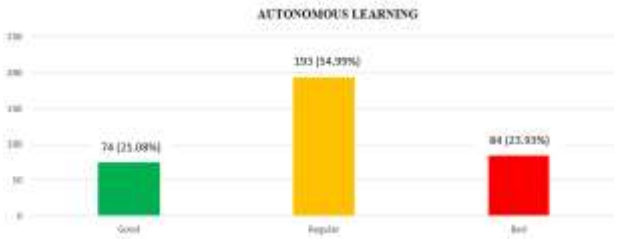

Fig. 3: Machine learning dimension level

Figure 4 shows the results of the knowledge creation dimension of the knowledge management variable. 41.88% of the contributors indicate that the level of knowledge creation is fair, suggesting that although organizations are generating knowledge, this process could be more effective. 31.05% of the employees consider the level poor, highlighting the need to improve strategies and methods for knowledge creation. On the other hand, 27.07% say

the level is good, indicating that some organizations are achieving good results, although there is still ample room for widespread improvement.

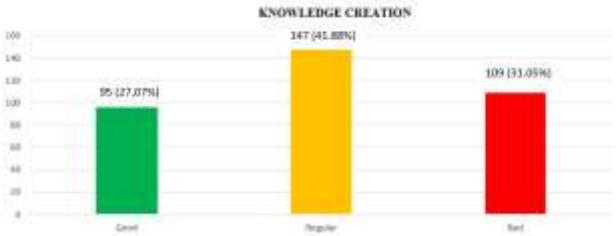

Fig. 4: Level of the knowledge creation dimension

Figure 5 shows the results of the knowledge storage dimension of the knowledge management variable, where a worrying 61.25% of employees indicate that the level of knowledge storage is poor. This result highlights a severe deficiency in the ability of organizations to store and organize knowledge effectively, which can lead to significant loss of valuable information and duplication of effort. Only 26.78% consider the level good, and 11.97% consider it fair, suggesting that current knowledge storage practices are primarily inadequate, and that significant improvement is needed to optimize knowledge management.

Fig. 5: Level of the knowledge storage dimension

4.1 Normality Test

Table 1 shows the results of the Kolmogorov-Smirnov and Shapiro-Wilk normality tests, indicating that all dimensions (data mining, predictive analytics, machine learning, knowledge creation, and knowledge storage) have a non-parametric distribution ($p < 0.05$ for all dimensions). Because of this, it is recommended to use non-parametric correlations, such as Spearman's or Kendall's correlation, instead of Pearson's correlation, to analyze the relationships between these variables, as non-parametric tests do not assume normality in the data.

Table 1. Normality test for Data Analysis and Knowledge Management dimensions

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Stadistic	gl	Sig.	Stadistic	gl	Sig.
Data extraction	,275	351	,000	,800	351	,000
Predictive analysis	,342	351	,000	,637	351	,000
Machine learning	,466	351	,000	,541	351	,000
Knowledge creation	,251	351	,000	,807	351	,000
Knowledge storage	,316	351	,000	,748	351	,000

a. Corrección de significación de Lilliefors

4.2 Spearman Correlation

As shown in Table 2, Spearman correlation results have been obtained that provide information on the relationships between the different dimensions analyzed in this study. A statistically significant positive correlation was found between the dimensions of data mining and knowledge creation (r=,889, p< .001). This indicates that as the levels of data extraction improve, workers report a better level of knowledge creation. Also, a very strong positive correlation was found between the predictive analytics and knowledge creation dimensions (r=,943, p< .001). This suggests that the better the predictive analytics, the better the knowledge creation.

Similarly, it was evident that there is a moderate positive correlation between data mining and knowledge storage dimensions (r=,498, p< .001). This indicates that as data extraction improves, workers report better data storage. Finally, a strong positive correlation was found between predictive analytics and knowledge storage (r=,706, p< .001). This means that as levels of predictive analytics improve, workers report improved knowledge storage.

Table 2. Relationship between the dimensions of the data analytics and work knowledge management variables

		1	2	3	4	5
Rho de Spearman	1. Data extraction	____	,942**	,543**	,889**	,498**
	2. Predictive analysis	,942**	____	,576**	,943**	,706**
	3. Machine learning	,543**	,576**	____	,815**	-0.003
	4. Knowledge creation	,889**	,943**	,815**	____	,500**
	5. Knowledge storage	,498**	,706**	-0.003	,500**	____

**. Correlation is significant at the 0.01 level (bilateral).

*. Correlation is significant at the 0.05 level (bilateral).

5 Proposal

Based on the survey results, the following data analytics model is proposed to improve knowledge management. This model allows evaluating the current state, applying the model, and obtaining suitable results.

Figure 6 presents a data analytics model designed to improve knowledge management in retail companies. It highlights the transition from the "Real State" to the "Ideal State" through a series of specific interventions.

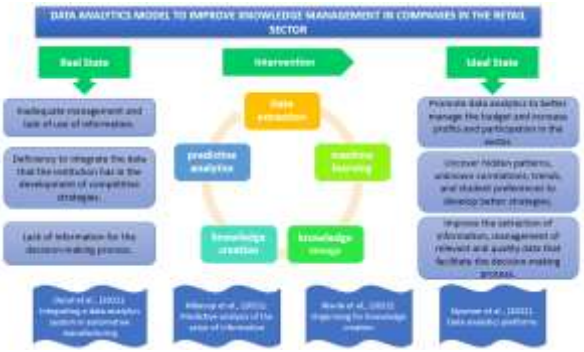

Fig. 6: Strategic proposal for improving data analytics

In the Real estate, retail companies face several significant challenges. Inadequate management and lack of exploitation of available information prevent organizations from fully utilizing their data. In addition, this data needs to be integrated more into developing competitive strategies, which limits their ability to stay ahead of the market. Finally, more relevant and valuable information must be provided to ensure the decision-making process, positively affecting organizational efficiency and effectiveness.

The Intervention proposed in the model includes three key components: data mining, predictive analytics, and machine learning. Data mining focuses on obtaining relevant information from various sources, ensuring the data is complete and valuable. Predictive analytics uses advanced techniques to forecast future trends and behaviors, allowing companies to anticipate market changes. Machine learning applies algorithms that allow machines to learn from data and improve their predictions and decisions over time. These components are designed to improve two critical dimensions of knowledge management: knowledge creation and knowledge storage. Knowledge creation involves generating new insights from analyzed data, while knowledge storage refers to storing and organizing this information efficiently for future use.

These interventions aim to reach an Ideal State where data analytics is a fundamental tool for budget management, profit growth, and industry participation. In this ideal state, companies can discover hidden patterns, unknown correlations, trends, and preferences, using these insights to develop better business strategies. In addition, information extraction and the management of relevant and quality data will be improved, facilitating informed and effective decision-making.

The contributions of this study are significant for both theory and practice in the field of knowledge management and data analytics in the retail sector. Theoretically, the proposed model provides a clear and structured framework that integrates data mining, predictive analytics, and machine learning, highlighting their impact on knowledge creation and storage. This provides a solid foundation for future research exploring or expanding these dimensions. At the practical level, the implications of this study are profound. Retail organizations can significantly improve their information management and decision-making by identifying deficiencies and proposing specific interventions. Implementing these practices can lead to better budget management, increased profits, and more significant market share, fostering an organizational culture based on accurate data and knowledge. In addition, by improving the ability to uncover hidden patterns and trends, companies can develop more effective and competitive strategies, strengthening their position in the industry.

6 Discussion

In Figure 1, 52.99% of collaborators indicate that the level of data extraction in the institution is poor, which shows that the institution has difficulties in collecting information from databases and repositories due to compatibility problems, data not structured, and of poor quality that impair the processing of information. This agrees with [14] and [16] who mentioned that data extraction is a flexible, efficient, and accessible process that brings together information from semi-structured and unstructured sources to speed up data analysis and reporting. Likewise, [15], states that it is necessary to work with specialized programs to improve data extraction through the construction of the sensor hardware, the development of the perception algorithm, and the scenario data.

In Figure 2, 57.83% of collaborators indicate that the level of predictive analysis in companies in the retail sector is poor, which indicates that staff must be trained to analyze and evaluate the data

handled by the company to anticipate the future. and discover new trends that generate positive results. Coinciding with [17], [19] who indicate that predictive analysis is an advanced study system that focuses on deeply examining a set of data, reports, and content to predict risks, opportunities, or behavior in the market. Likewise, [18] maintains that currently statistical modeling techniques and new technologies such as big data and machine learning are tools used by predictive analysis to make predictions that promote continuous improvement and improve flow management. information.

In Figure 3, 54.99% of collaborators maintain that the level of machine learning in the institution is regular, showing that the institution has qualified teams for the development of basic data collection and organization tasks. However, it is necessary to implement more complex programs to improve the response to tasks not explicitly programmed and the adaptability of the equipment according to the data stored in the system. This coincides with [20], [23] who mention that machine learning is part of artificial intelligence influencing the ability to learn and act of a computer in relation to the needs that are present and the information available for the machine to elaborate a new answer. In addition, [21], adds that machine learning is carried out by means of computer programs that use the information of the system and the patterns present in the data to carry out unprogrammed actions.

In Figure 4, 41.88% of collaborators affirm that the level of knowledge creation is regular, which shows that there is a strong intention to exchange data and valuable information among the members of the institution. However, there are difficulties in the data transfer and processing channels, which limits the ability to innovate and include new ideas in the organization's strategies. This agrees with [24], [28] who mention that the creation of knowledge is a collective process that includes the participants in a space to exchange and integrate all the knowledge they possess with the objective of formulating new ideas, which allows us to overcome the competition through constant innovation. Similarly, [26] points out that the creation of knowledge is essential to achieve continuous improvement, increase competitiveness in the market, and generate innovative strategies to meet the needs of the organization and customers.

In Figure 5, 61.25% of collaborators mention that the level of knowledge storage is poor, demonstrating that there are difficulties in retaining, organizing, and distributing knowledge in the institution, limiting the ability to consolidate knowledge and increase the interaction of members with valuable information held by the organization. Coinciding with [29], [31] who indicate that knowledge storage consists of organizing and distributing the knowledge of the organization in various databases, intranets, extranets, and information systems that improve the provision of information to innovate, lead, and direct strategies. Likewise, [30] affirms that the storage of knowledge allows to safeguard the new theories, patterns, ideas, and information generated by the organization, which facilitates the interaction of workers with the information to increase productivity levels.

7 Conclusions

This research proposal maintains that data analytics is essential to improving data collection, evaluation, and analysis, as well as finding patterns, correlations, and trends that improve the organization's strategies. This increases the capacity for innovation and decision-making based on the new knowledge generated.

Likewise, machine learning allows simple tasks related to data analytics to respond to the organization's needs. However, complex programs that improve the ability to react to unscheduled tasks through efficient information management must be implemented. Also, information can adapt data patterns automatically in the responses provided by technological equipment in unknown situations. Similarly, creating knowledge actively promotes the implementation of spaces for exchanging essential data and information. However, the knowledge distribution and analysis network must be improved and automated to integrate them into organizational strategies.

On the other hand, data extraction is deficient since the process of collecting information from databases faces incompatibility problems, unstructured or semi-structured data, poor data quality, and data security, which limits the ability of computers to recognize, collect, analyze, process, and organize information. In addition, predictive analytics cannot make safe and accurate predictions based on available data, so the information tends to be distorted, inadequate, or difficult to understand. Finally, knowledge storage is inadequate because no secure methods exist to retain, organize, and distribute the information generated. This limits the organization's ability to consolidate the knowledge produced, share the information, and include it in developing strategies, decision-making, and organizational plans.

8 Limitations, Recommendations and Future Work

While providing a valuable assessment of the current state of data analytics in the retail sector, this study has certain limitations. First, the research was based on surveys only, which may introduce self-reporting biases and limit the depth of insights obtained. In addition, the sample is restricted to employees of companies in the retail sector, which may be different from other sectors or have a broader view of the problem. Future studies should incorporate mixed methodologies that include qualitative and quantitative approaches and expand the sample to include different sectors and hierarchical levels within organizations. In addition, it is suggested that longitudinal studies be implemented to observe the evolution and impact of data analytics over time. Finally, future research could further explore the barriers organizations face in implementing advanced data analytics technologies and how these can be overcome through training and organizational change strategies.

Acknowledgement:
We are grateful to the Universidad San Ignacio de Loyola for their support during the conduct of this research. Also, the authors would like to acknowledge the financial support of the National Funds provided by FCT-Foundation for Science and Technology to VALORIZA-Research Center for Endogenous Resource Valorization (project UIDB/05064/2020).

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used Chat GTP in order to briefly clarify joint concepts. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References:

[1] M. Pohl, D. Staegemann, and K. Turowski, "The Performance Benefit of Data Analytics Applications," *Procedia Comput Sci,* vol. 201, pp. 679–683, Jan. 2022, doi: 10.1016/J.PROCS.2022.03.090.

[2] N. Nyoman, B. Tekinerdogan, C. Catal, and R. Tol, "Data analytics platforms for agricultural systems: A systematic literature review," *Comput Electron Agric,* vol. 195, p. 106813, Apr. 2022, doi: 10.1016/J.COMPAG.2022.106813.

[3] A. Geistanger, K. Braese, and R. Laubender, "Automated data analytics workflow for stability experiments based on regression analysis," *Journal of Mass Spectrometry and Advances in the Clinical Lab*, vol. 24, pp. 5–14, Apr. 2022, doi: 10.1016/J.JMSACL.2022.01.001.

[4] A. Perdana, H. Lee, D. Arisandi, and S. Koh, "Accelerating data analytics adoption in small and mid-size enterprises: A Singapore context," *Technol Soc*, vol. 69, p. 101966, May 2022, doi: 10.1016/J.TECHSOC.2022.101966.

[5] A. Dacal, J. Areal, V. Alonso, and M. Lluch, "Integrating a data analytics system in automotive manufacturing: background, methodology and learned lessons," *Procedia Comput Sci,* vol. 200, pp. 718–726, Jan. 2022, doi: 10.1016/J.PROCS.2022.01.270.

[6] W. Sardjono, Harisno, and W. G. Perdana, "Improve Understanding and Dissemination of Disaster Management and Climate Change by Using Knowledge Management Systems," *IOP Conf Ser Earth Environ Sci,* vol. 426, no. 1, p. 012158, Feb. 2020, doi: 10.1088/1755-1315/426/1/012158.

[7] O. Barbón and J. Fernández, Role of strategic educational management in the management of knowledge, science, technology and innovation in higher education ("Rol de la gestión educativa estratégica en la gestión del conocimiento, la ciencia, la tecnología y la innovación en la educación superior"), *Educación Médica*, vol. 19, no. 1, pp. 51–55, Jan. 2018, doi: 10.1016/J.EDUMED.2016.12.001.

[8] M. Mohammad, R. Abdullah, A. Jabar, W. Sardjono, M. Mukhlis, and E. Selviyanti, "Factors to increasing the employee performance through knowledge management systems implementation at PT. XYZ," *J Phys Conf Ser*, vol. 1563, no. 1, p. 012023, Jun. 2020, doi: 10.1088/1742-6596/1563/1/012023.

[9] R. Abdul, D. Maria, S. Laila, and M. Azima, "Development of Knowledge Management System for Determining Organizational Performances, Total Quality Management, And Culture," *J Phys Conf Ser,* vol. 1529, no. 2, p. 022063, Apr. 2020, doi: 10.1088/1742-6596/1529/2/022063.

[10] G. Li and L. Jiajun, "Automatic Analysis And Intelligent Information Extraction Of Remote Sensing Big Data," *J Phys Conf Ser*, vol. 1616, no. 1, p. 012003, Aug. 2020, doi: 10.1088/1742-6596/1616/1/012003.

[11] S. Yusoff, N. Noh, and N. Isa, "University Students' Readiness for Job Opportunities in Big Data Analytics," *J Phys Conf Ser*, vol. 2084, no. 1, p. 012026, Nov. 2021, doi: 10.1088/1742-6596/2084/1/012026.

[12] R. Rawat and R. Yadav, "Big Data: Big Data Analysis, Issues and Challenges and Technologies," *IOP Conf Ser Mater Sci Eng*, vol. 1022, no. 1, p. 012014, Jan. 2021, doi: 10.1088/1757-899X/1022/1/012014.

[13] A. Rojas, J. Londoño, N. Pérez, and M. Gómez, "Analysis of the big data generated in the company's social networks 'Sistemas Expertos SAS' using NVivo," *J Phys Conf Ser,* vol. 1418, no. 1, p. 012004, Dec. 2019, doi: 10.1088/1742-6596/1418/1/012004.

[14] S. Hu and H. Yin, "Research on the optimum synchronous network search data extraction based on swarm intelligence algorithm," *Future Generation Computer Systems*, vol. 125, pp. 151–155, Dec. 2021, doi: 10.1016/J.FUTURE.2021.05.001.

[15] A. Razak, S. Asmah, L. Wang, B. Yu, and C. Chen, "Parking Area Data Collection and Scenario Extraction for the Purpose of Automatic Parking ADAS Function," *IOP Conf Ser Mater Sci Eng*, vol. 780, no. 3, p. 032026, Mar. 2020, doi: 10.1088/1757-899X/780/3/032026.

[16] J. Nolde, A. Mian, L. Schlaich,, J. Chan, L. Lugo-Gavidia, N. Barrie, V. Gopal, G. Hillis, C. Chow and M. Schlaich. "Automatic data extraction from 24 hour blood pressure measurement reports of a large multicenter clinical trial," *Comput Methods Programs Biomed*, vol. 214, p. 106588, Feb. 2022, doi: 10.1016/J.CMPB.2021.106588.

[17] A. Tolba and Z. Al, "Predictive data analysis approach for securing medical data in smart grid healthcare systems," *Future Generation Computer Systems,* vol. 117, pp. 87–96, Apr. 2021, doi: 10.1016/J.FUTURE.2020.11.008.

[18] A. Mbakop, F. Biyeme, J. Voufo, and J. Lucien, "Predictive analysis of the value of information flow on the shop floor of developing countries using artificial neural network based deep learning," *Heliyon,* vol. 7, no. 11, p. e08315, Nov. 2021, doi: 10.1016/J.HELIYON.2021.E08315.

[19] T. Drozdova and A. Vereshchagina, "Predictive assessment of man-made risks during oil-handling operations at tank farms," *IOP Conf Ser Earth Environ Sci,* vol. 408, no. 1, p. 012017, Jan. 2020, doi: 10.1088/1755-1315/408/1/012017.

[20] K. Jayareka, P. Sobiyaa, Kaladevi, Vinodhini.V, and B. Suman, "An effective automatic detection of tooth cavity using machine cum deep learning concepts and ICDAS measurement," *Mater Today Proc,* May 2022, doi: 10.1016/J.MATPR.2022.05.109.

[21] M. Su, B. Liang, S. Ma, C. Xiang, C. Zhang, and J. Wang, "Automatic Machine Learning Method for Hyper-parameter Search," *J Phys Conf Ser*, vol. 1802, no. 3, p. 032082, Mar. 2021, doi: 10.1088/1742-6596/1802/3/032082.

[22] R. Castilla, A. Pacheco, I. Robles, A. Reyes, and R. Inquilla, "Digital channel for interaction with citizens in public sector entities," *World Journal of Engineering*, vol. 18, no. 4, pp. 547–552, 2020, doi: 10.1108/WJE-08-2020-0377. .

[23] J. Qu and X. Cui, "Automatic machine learning Framework for Forest fire forecasting," *J Phys Conf Ser*, vol. 1651, no. 1, p. 012116, Nov. 2020, doi: 10.1088/1742-6596/1651/1/012116.

[24] A. Nisula, K. Blomqvist, J. Bergman, and S. Yrjölä, "Organizing for knowledge creation in a strategic interorganizational innovation project," *International Journal of Project Management,* vol. 40, no. 4, pp. 398–410, May 2022, doi: 10.1016/J.IJPROMAN.2022.03.011.

[25] J. Ore, A. Pacheco, E. Roque, A. Reyes, and L. Pacheco, "Augmented reality for the treatment of arachnophobia: exposure therapy," *World Journal of Engineering*, vol. 18, no. 4, pp. 566–572, 2020, doi: 10.1108/WJE-09-2020-0410. .

[26] K. Al, D. Palacios, and K. Ulrich, "The impact of intellectual capital on supply chain agility and collaborative knowledge creation in responding to unprecedented pandemic crises," *Technol Forecast Soc Change,* vol. 178, p. 121603, May 2022, doi: 10.1016/J.TECHFORE.2022.121603.

[27] R. Castilla, A. Pacheco, and J. Franco, "Digital government: Mobile applications and their impact on access to public information," *SoftwareX,* vol. 22, p. 101382,

May 2023, doi: 10.1016/J.SOFTX.2023.101382.

[28] S. Juvonen, J. Koivisto, and H. Toiviainen, "Knowledge creation for the future of integrated health and social services: Vague visions or an expansion of activity?," *Learn Cult Soc Interact*, p. 100613, Feb. 2022, doi: 10.1016/J.LCSI.2022.100613.

[29] C. Mubin and Y. Latief, "Organizational culture influence on implementation of knowledge management and quality management system for improving Indonesian construction companies' performances," *IOP Conf Ser Mater Sci Eng*, vol. 508, no. 1, p. 012037, Apr. 2019, doi: 10.1088/1757-899X/508/1/012037.

[30] M. Alim, A. Nugroho, S. Arif, Murtiningrum, and L. Sutiarso, "Development of knowledge management system for assisting the Agrotechno Edu-park establishments in Sriharjo village, Imogiri district, Bantul regency," *IOP Conf Ser Earth Environ Sci*, vol. 542, no. 1, p. 012064, Jul. 2020, doi: 10.1088/1755-1315/542/1/012064.

[31] P. Chaithanapat, P. Punnakitikashem, N. C. Khin Khin Oo, and S. Rakthin, "Relationships among knowledge-oriented leadership, customer knowledge management, innovation quality and firm performance in SMEs," *Journal of Innovation & Knowledge*, vol. 7, no. 1, p. 100162, Jan. 2022, doi: 10.1016/J.JIK.2022.100162.

[32] R. Hernández and C. Mendoza, *Research methodology: the three quantitative, qualitative and mixed routes (Metodología de la investigación: las tres rutas cuantitativa, cualitativa y mixta)*. México: Mc Graw Hill- educación, 2018.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

- Rosario Pariona-Luque and Alex Pacheco were in charge of writing and revising the article.
- Edwin Vegas-Gallo, Edwin Felix-Poicon and Rui Alexandre Castanho, Ana Loures, and Liz Pacheco-Pumaleque carried out the conceptualisation and methodology of the research.
- Fabian Lema collected and curated the data.
- Marco Añaños-Bedriñana and Wilson Marin conducted the research.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare.