CIGI QUALITA MOSIM 2023

Machine Learning for Decision Making in the Improvement of Digital Press Services

EYA KALBOUSSI, NADIA NDHAIEF, NIDHAL REZG

Laboratoire de Génie Informatique de Production et de Maintenance (LGIPM), Université de Lorraine, France
eya.kalboussi@univ-lorraine.fr, nadia.ndhaief@univ-lorraine.fr, nidhal.rezg@univ-lorraine.fr

***Résumé* – Suite aux crises actuelles économiques, environnementales et sanitaires, les entreprises doivent être plus flexibles et réactives afin de survivre. Une information de haute qualité et d'une fidélité robuste, est très primordiale. Elle sert à améliorer les processus de prise de décision dans les entreprises afin de garantir leur succès. Dans cet article, un processus de prise de décision stratégique à base d'apprentissage automatique appliqué à un cas réel au sein du Group6TM, qui est une entreprise ouvrant dans le développement du contenu audiovisuel utiles aux décideurs dans les entreprises de divers secteurs pour la prise des décisions stratégiques et l'amélioration des Indicateurs Clés de Performance (KPIs). Notre approche proposée dans ce papier se base d'abord sur la construction de tableaux de bord présentant une vue claire des données utiles permettant une prise de décision rapide et par les décideurs afin de mieux comprendre nos données et améliorer notre approche, la deuxième partie du travail est consacrée à l'utilisation d'un algorithme simplifié basé sur la technique d'apprentissage automatique permettant un déploiement rapide pour les entreprises qui offrent de supports de communication, de services informatiques et d'événements afin de soutenir le développement des industries professionnelles et en conséquence, faciliter la prise de décision quotidienne des professionnels. Cet algorithme a pour objectif d'améliorer les prises de décisions au sein des entreprises afin de maximiser ses KPIs.**

***Abstract* – Owing to the current economic, environmental and health crises, companies must be more flexible and reactive in order to survive. High-quality information with robust fidelity is crucial as it serves to improve their decision-making processes and leads to their success. In this article, a strategic decision-making approach process based on Machine Learning (ML) technique including a real case study within Group6TM, which is a company opening in the development of audio-visual content useful to decision makers in companies of various sectors for making strategic decisions and improving Key Performance Indicators (KPIs). Our proposed approach is first based on building dashboards that present data in a clear way and speed up decision-makers' ability to make decisions. To get powerful insights from data and improve our approach, the second part of the work is devoted to the use of a simplified algorithm allowing a quick deployment for any company that offers service communication media, information technology services and events in order to support the development of professional industries and consequently, facilitate the daily decision making of professionals. This algorithm aims to improve decision making within companies in order to maximize its KPIs.**

***Mots clés* – Analyse des données, Tableaux de bord, Power BI, Apprentissage automatique.**
***Keywords* – Data Analysis, Dashboards, Power BI, Machine Learning.**

1 INTRODUCTION

Due to the changing health climate caused by the Covid-19 pandemic, fierce competition and ongoing wars, some sectors have slowed down their activities. These factors forced companies to be more flexible and resilient in the events of crises. therefore, the use of Artificial Intelligence (AI) tools that enables them to make decisions effectively and rapidly [Pietronudo and al., 2022]. Moreover, the rapid expansion of data opens the door for the use of the most advanced techniques of data science [Xu and al., 2021] that are revolutionizing the world and opening up new horizons, such as Machine Learning (ML), Deep Learning (DL)-and Industrial Internet of Things in all fields across all industries to make better and faster decisions and deal with the vagaries of life [Wittenberg, 2022].
By mastering data analysis and diverse data science techniques, it is possible to create new innovative products and good services as well as help improve business decision-making.
By referring to the literature, researchers have built extensive expertise over the years that covers most data analysis challenges in a variety of industries, such as engineering, medicine, economics, and business management as well as marketing [Woschnak and al., 2020]. However, our study is based on the use of ML and DL in the context of urban logistics, audio-visual and digital press services.
In order to understand and point out the current state of the research and provide future insights, this paper focuses on reviewing the existing issues on these emerging topics.
First, we begin by illustrating the literature reviews conducted in the aforementioned domain.
We found recent relevant articles on logistics in which ML techniques were used to solve the given problems. ML and data mining were used to solve freight transportation and supply chain issues by [Tsolaki and al., 2022]. According to Qu and Li [2022], the ant colony algorithm was employed to achieve optimization of terminal distribution. Dossou and Vermersch [2021] generated a decision-aided tool for sustainable urban logistics optimization to make circulation more fluid. Shi and al., [2019] developed a qualitative and quantitative method that combines an online intelligent scheduling approach with AI technology. A self-adaptative micro-level control by combining reinforcement learning and rule-based agent models was

proposed by Bosse [2020]. The research of [Singh and al., 2021] is significant for its numerous recent and pertinent articles about the use of supervised, unsupervised and reinforcement ML techniques to solve logistics problems. Based on the findings of a systematic literature review, Woschank and al., [2020] claimed that using ML and DL ensure smart logistics and smart manufacturing. Giuffrida and al., [2022] discussed last-mile logistics optimization with the help of -ML approaches, and they conducted a bibliometric analysis and a critical review to underline the major issues, algorithms used, and case studies.

Regarding the audio-visual football issues, researchers have used a data science approach to minimize the time required to select players for a team by considering cost and player skill as constraints [Rajesh and al., 2020]. Andriyanov [2020] built data clustering algorithms based on Gaussian Mixture Model (GMM) and neural networks algorithms and did a comparative analysis of the accuracy of clustering algorithms which produce the same result. However, the GMM appears to be more accurate. Zhao and Dong [2022] examined player interaction in a football video using scene analysis and other methods. Researchers have used a variety of algorithms ranging from Rodriguesa and Pinto [2022], who tested various ML methods such as SVM, RF, Xgboost that use multiple statistics from previous matches and attributes of players from the two teams as input for predicting the outcome of a football match. The best model was RF that has the best accuracy of 25,26%. Anfilets and al., [2019] who used a deep multilayered neural network for predicting the match result with an accuracy of 61.14%. Rathi and al., [2020] have illustrated the potential of AI and its limitations in football. The advantages of AI include improved decision-making, player selection process and ticket price prediction. However, it has some limitations like high cost, absence of feelings and emotions, increasing dependency on machines, lack of both original creativity and thinking out of the box. Majumdar and al., [2022] investigated injury prediction in football to help understand adaptability to training programs, evaluate exhaustion and recovery, and reduce the risk of injury and illness.

Concerning social media, Ali and al., [2023] proposed a Long - short-term memory-Gated Recurrent Unit (LSTM-GRU) model to classify hate content into categories and identify influential people’s participation in the virtual communities with an accuracy of 98.14%. A ML tool called Waikato Environment for Knowledge Analysis (WEKA) to predict and identify online consumer's behavior in order to create an effective marketing strategy is proposed in [Arasu and al., 2020]. Li and al., [2022] applied a self-learning semi-supervised DL network that trains supervised and unsupervised tasks at the same time to detect fake news on social media and provides a precision of 0.9. Babu and Kanaga [2021] proposed various AI techniques for the sentiment analysis for depression detection, including ML techniques, DL techniques and social media analysis, and a higher accuracy is obtained with DL. Sharma and Shafiq [2021] researched on intentions of their customers or users through different channels like platforms, posts and discussions on social media. Their solution carried out different tools such as ML and DL models to determine and predict the clients purchase intentions. Light Gradient Boosting Framework (LGBM) algorithm produces superior results to others. As mentioned in [Sufi and Alsulami, 2022], an approach utilized AI techniques, language detection, translation, sentiment analysis, and named entity recognition (NER) to collect Geospatial intelligence information from social media posts of political leaders has achieved an accuracy of 97%. Majewska and Majewska [2022] showed through their researches the vital role of social media in maximizing football clubs’ revenues using analysis techniques.

In previous works, several techniques that were used to enhance logistics, football and digital performance are classified into four groups as illustrated in the summary Table 1 of previous studies which displays an overview of the AI techniques employed in the context of improving the performance of fields mentioned above. The most commonly used technique of AI is ML which will be used in our case.

In that Regard, our work aims to solve the biggest problems in digital press services with the new digital technologies such as AI, big data and analytics. Thereafter, we will adapt our decision-making tool support specifically for the logistics and audio-visual industries, with the goal of improving performance and enabling decision-makers to make more informed and effective decisions in these domains.

We worked with Group6TM, a company specialized in digital media, to propose a strategic decision-making process based on an "intelligent algorithm" as unsupervised learning.

Hence, the purpose of this paper is to develop a decision support process for this company, considered as a leader in several B-to-B markets in France and seeks to manage its data flow. Bringing relevant information to decision makers helps them in making strategic decision based on the most used techniques of AI.

This group focuses on developing a judicious solution in order to contribute to the development of professional sectors (logistics, manufacturing, banks, insurance, etc.), by facilitating their daily decisions with accurate and useful information and a wide range of services.

Table 1. Overview of the AI techniques employed in logistics, football and digital press fields.

		[Tsolaki and al., 2022]	[Qu and Li, 2022]	[Dossou and Vermersch, 2021]	[Shi and al., 2019]	[Bosse, 2020]	[Singh and al., 2021]	[Woschnak and al., 2020]	Giuffrida and al., [2022]	[Rajesh and al., 2020]	[Andriyanov, 2020]	[Zhao and Don, 2022]	[Rodriguesa and Pinto, 2022]	[Anfilets and al., 2019]	[Rathi and al., 2020]	[Majumdar and al., 2022]	[Ali and al., 2023]	[Li and al., 2022]	[Babu and Kanaga, 2021]	[Sharma and Shafiq, 2021]	[Sufi and Alsulami, 2022]	Total
AI techniques	ML	x		x	x	x	x	x	x	x		x	x		x	x		x	x	x	x	16
	DL	x		x			x	x			x			x			x		x	x		9
	Optimization metaheuristics: The ant colony		x																			1
	Multi-agent system			x		x	x															3
Field services	Logistics	x	x	x	x	x	x	x	x													8
	Audio-visual									x	x	x	x	x	x	x						7
	Digital press																x	x	x	x	x	5

The rest of the paper is structured as follows. For the second section, we specify the detailed decision-making process of our proposed method. Section 3 elaborates the methodology used and describes each approach adopted in the proposed model. We were interested in building dashboards using different dashboard software and using ML algorithm to get more insights from our data The paper ends with a conclusion and also mentions the directions for further research.

2 PROPOSED METHOD

The proposed system for performing decision-making which can be applied in differents sectors is provided in Figure 1. The involved processes are represented using the blocks of the diagram to build an efficient decision support tool. It is well known that information is key. Accurate, reliable and timely information is essential to improve decision-making. For that reason, it was necessary to create dashboards that are helpful To effectively monitor and analyze the data. In addition, it helps to quickly detect issues that requiers solving. In this article, our work will be divided into two parts as follows:

- Building dashboards to meet the goal of the company.
- Implementing a clustering algorithm K-Means for separating data based on their similarities.

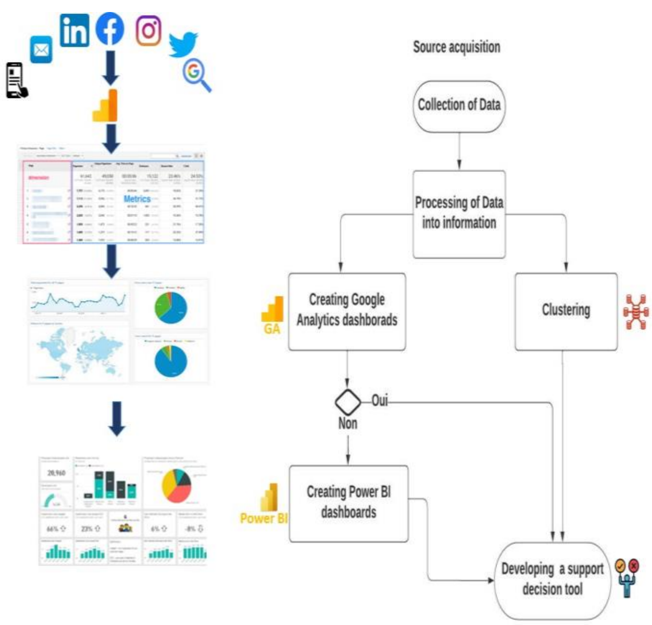

Figure 1. Flowchart of the proposed method

As presented in the previous figure, the first part consists on collecting data through GA. This acquires data from multiple traffic sources including Facebook, Instagram, E-mails and twitter to provide general track information. It includes pageviews, device type, acquisition channels, operating system, sessions, average session duration, new users, users, bounce rate, pages views per session, top pages views, average pages per session, user location, and user language. In the second step, we will create GA dashboards that presents an overview of the most important metrics mentioned above. Considering the case where the dashboards are not very interactive, we will use Power BI to build more creative ones. Thereafter, it’s a matter of employing a clustering algorithm to get better ideas about our data and its structure. Finally, our decision support tool will be ready and reliable to improve the company's performance.

3 METHOD IMPLEMENTATION

In this section, we will present our purpose by presenting our study's results which focuses on a real case of a media group Group6TM. This group is specialized in information communication and professional press. Its goal is to wisely manage their social media, by monitoring their followers count and tracking their global presence and activities. This section describes how we collected data, how we analyze the data using dashboards and ML model for clustering all data collected.

3.1 Dashboards for Data Analysis

In this subsection, the considered dashboards for data analysis aiming to track and analyze audience data are described.

3.1.1 Google Analytics (GA) dashboards

This subsection is dedicated to discussing the GA dashboards. The primary reason for utilizing Google Analytics in this study is its ability to provide time series data. Additionally, it is a free service offered by Google that generates detailed statistics on website visits. The platform enables website owners to track the source of the website's traffic, including search engines, referral sites, emails, and direct visits, providing valuable insights on how to enhance the website's content and design [Plaza, 2011]. Hence, in Figure 2 below, we will illustrate the GA audience dashboard which shows several metrics that are useful for analyzing the audience. The ones displayed are the number of users, new users, sessions, users, bounce rate, the evolution of sessions, traffic source. And adding to that, the most used social networks sites and apps to visit and the evolution of users during a week, all that so we can be able to determine the peak period during which we have more active sessions and users. As well, the most viewed pages that generate the most traffic.

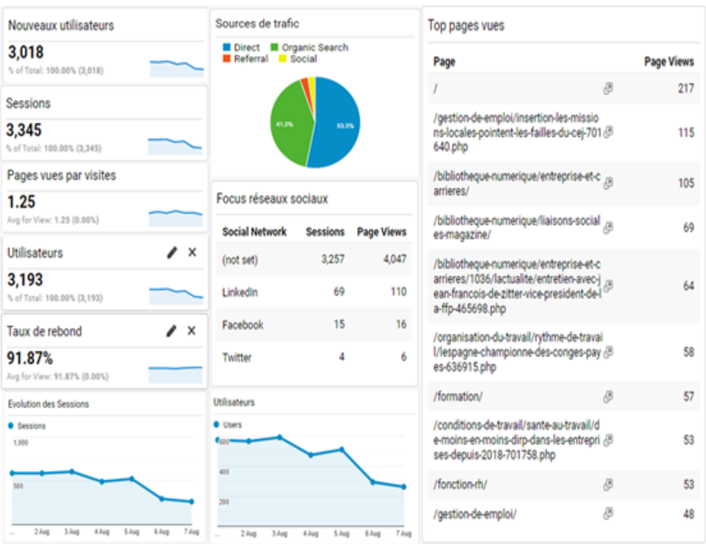

Figure 2. GA Dashboard

3.1.2 Power BI dashboards

Thereafter, we chose to use the Power BI software, data visualizations tool, because it offers the possibility of building more interactive and smart dashboards. First, the historical data was imported into Power BI using the GA connector as shown in Figure 3. Second, we employed Data Analysis Expressions (DAX) in order to create smarter measures by which we built an effective Power BI dashboard. DAX is another important advantage of Power BI, with a library that allows users to be able to use operators and functions to create new measures in Power BI [Sousa and al., 2021] as represented in Table 2 below that illustrates some examples of DAX expressions used to create new measures.

Table 2. DAX expressions

Measures	DAX Formulas
Gross evolution of pages views	(SUM('Session'[Pageviews]) - [Pageviews N-1]) / [Pageviews N-1]
New users of the year N-1	CALCULATE (SUM ('Session' [New Users]), ALL('Date'), USERELATIONSHIP('Date'[Date], 'Previous Date'[Date])
Bounces of the year N-1	CALCULATE (SUM('Session'[Bounces]), ALL ('Date'), USERELATIONSHIP ('Date'[Date], 'Previous Date'[Date])
Evolution of the average session duration	FORMAT (((SUM ('Session' [Session Duration]) /SUM('Session'[Sessions])) - ([Session Duration N-1] / [Sessions N-1])) / ([Session Duration N-1] / [Sessions N-1])
Pages views/session	DIVIDE(SUM('Session'[Pageviews]), SUM('Session'[Sessions])
Evolution of the bounce rate	FORMAT(((sum('Session'[Bounces])/ sum('Session'[Sessions])) -([Bounces N-1] / [Sessions N-1])) / ([Bounces N-1] / [Sessions N-1])

After finishing creating the new measures, we presented an interactive dashboard for audience analysis filtered by date to make comparisons with the above. The Figure 3 shows audience dashboard that tracks the number of sessions per device and per channel acquisition in a determined period of time using a date filter. We created others dashboards to follow the evolution of others metrics between the year (N) and the year (N-1).

Figure 3. Power BI Dashboard

3.2 *Data Analysis using ML*

As explained in the previous subsection, we have presented audience analysis through the development of dashboards that focus on key metrics. We have also highlighted the utility of each one that has already been constructed. Thus, to fully understand the structure of our data, we will use the clustering technique in order to generate groups with high similarities. In this subsection, we will discuss our proposed process of analysis based on the ML technique conducted on data that was collected from GA by the supermetrics tool.

Figure 4 illustrates the analysis process of data using ML that will be developed below.

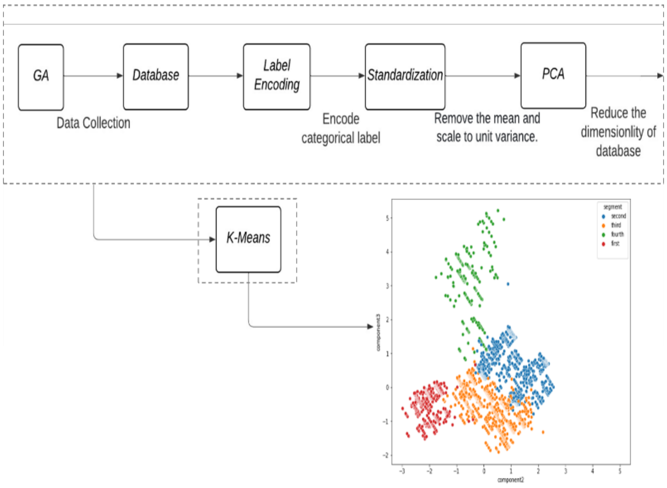

Figure 4. Analysis process of Data using ML

As already mentioned, the used data in this work was collected from GA using supermetrics which is the easiest way to transfer the marketing data to any format [Nanda, 2022] by running some queries. In this case, the format used was Comma-Separated Values (CSV). The dataset consists of 19104 press titles. Every press has 11 features including view, year, month, social network, country, user type, page title, device category, users, new users and unique pageviews. Once the data was collected, it was prepared for clustering according to the following order. The first step consists on Label Encoding, then standardization and finally Principal Component Analysis (PCA). The sequence of tools which were carried out for clustering is described in figure 5 below.

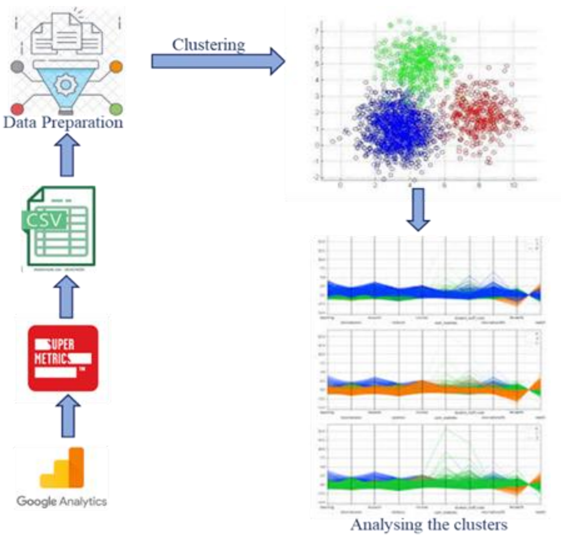

Figure 5. Sequence of used tools for clustering data

Label encoding approach converts the categorical variables into numeric. With this, dataset becomes numeric but includes features with different ranges. Therefore, standardization was applied to normalize our data. Then, the PCA, which is a dimensionality-reduction method, was applied on the standardized data to reduce the dimensionality of features. Hence, the number of features kept based on the rule of thumb is 7 components from 11 components. After further processing, clustering algorithm K-Means was applied to divide our data according to their principal characteristics into different groups is called cluster [Sameera and Shashi, 2022]. The clustering algorithm used was K-Means, it is the most familiar and popular unsupervised ML algorithm and the most often used [Barradas and al., 2022]. So, the number of clusters determined is 4 using the Elbow method, as shown in Figure 5.

The clustering algorithm did not provide useful details on the similarity among data points. Consequently, we proposed a parallel coordinates plot in order to study the characteristics of each group illustrated above. This plot was used to see the relationship between individual data points and all the variables in the dataset. In other words, how data points of each cluster

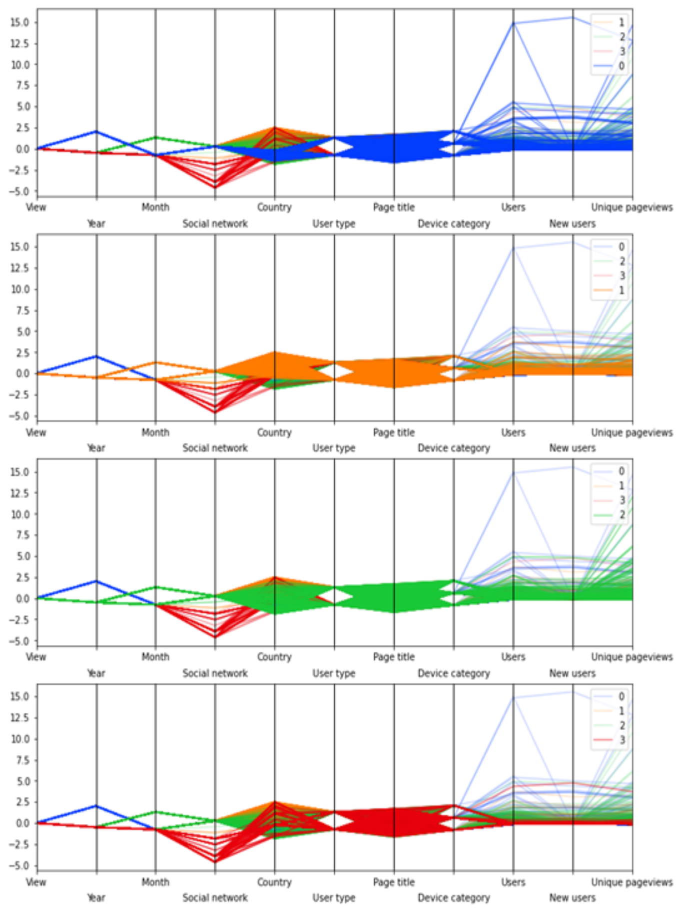

Figure 6. Parallel Coordinates Plot for the clusters

are distributed across all features as illustrated in figure 6 below. In the chart, the colors indicate distinct clusters. By analyzing how the values for each variable differ among the clusters, we can gain insights into the nature of each cluster.

The Figure above clearly illustrates the pertinent features that define each cluster obtained by K-Means algorithm. In fact, each color represents a different cluster. The blue samples are the cluster 0, which is characterized by a high proportion of unique page views, and a high number of users and new users. Cluster 1, whose samples are in orange is distinguished by high proportion users from the same country using the same device acquisition. Some of the green samples that represents cluster 2 have a high number of unique page views and a high proportion of users who use the same device acquisition. The red samples of cluster 3 are identified by a high proportion of users from the same country and a small proportion of active users of social media.

The following table 3 depicts the meaning of K-Means clusters to get a sense of clusters characteristics.

Table 3. Clusters interpretation

Cluster number	Cluster color	Cluster characteristics
0	Blue	- High proportion of unique page views. - High number of users and new users.
1	Orange	- High proportion users from the same country using the same device acquisition.
2	Green	- High number of unique page views. - High proportion of users who use the same device acquisition.
3	Red	- High proportion user from the same country. - Small proportion active user of social media.

4 CONCLUSION

To conclude, this work has allowed us to achieve several objectives through the decision support tool, including speed and efficiency. Nowadays, the process of decision-making is becoming more and more automated, which ensures an increase in productivity and a decrease in costs, justifying the need to propose better solutions based on the decision support tool. A case study of the company "Group6TM" via an optimal decision support tool provides the marketers of the group with interesting information to be more efficient. The aim of the proposed work is to control the information flow based on a decision tool support. To do this, we created dashboards that provide a complete analysis to marketers in order to manage the important social media marketing KPIs. These are the number of sessions, bounce rate, pages views per session and average session duration, which are gathered from various data sources, such as Facebook, Instagram and Twitter and all other social media platforms. Furthermore, dashboards help marketers focus on tracking users across marketing channels and devices. In addition, these dashboards provide future insights in order to help the Group6TM to make more effective and efficient strategic decisions. To highlight our solution, clustering algorithm was applied, aiming to ensemble our data according to their similarities into groups of similar data points into a sub-data named cluster. In fact, we exported our data from GA via supermetrics to an excel file. Next to that, this data was prepared for clustering in such a way we obtained four clusters in which the objects in the same cluster are similar.

For the continuation of our work, we will try to adapt our decision tool support for logistics and audio-visual field to help lead to performance enhancements and decision makers in making better decisions in these sectors.

5 REFERENCES

Pietronudo, M. C., Croidieu, G., & Schiavone, F. (2022). A solution looking for problems? A systematic literature review the rationalizing influence of artificial intelligence on decision-making in innovation management. *Technological Forecasting and Social Change*, *182*, 121828.

Xu, Z., Tang, N., Xu, C., & Cheng, X. (2021). Data science: connotation, methods, technologies, and development. *Data Science and Management*, *1*(1), 32-37.

Wittenberg, C. (2022). Challenges for the human-machine interaction in times of digitization, CPS & IIoT, and artificial intelligence in production systems. *IFAC-PapersOnLine*, *55*(29), 114-119.

Woschank, M., Rauch, E., & Zsifkovits, H. (2020). A review of further directions for artificial intelligence, machine learning, and deep learning in smart logistics. *Sustainability*, *12*(9), 3760.

Tsolaki, K., Vafeiadis, T., Nizamis, A., Ioannidis, D., & Tzovaras, D. (2022). Utilizing machine learning on freight transportation and logistics applications: a review. *ICT Express*.

Qu, L., & Li, H. (2022). Analysis of distribution path optimization algorithm based on big data technology. *Journal of King Saud University-Science*, 102019.

Dossou, P. E., & Vermersch, A. (2021). Development of a decision support tool for sustainable urban logistics optimization. *Procedia Computer Science*, *184*, 476-483.

Shi, H., Sun, L., Teng, Y., & Hu, X. (2019). An online intelligent vehicle routing and scheduling approach for B2C e-commerce urban logistics distribution. *Procedia Computer Science*, *159*, 2533-2542.

Bosse, S. (2020). Self-adaptive traffic and logistics flow control using learning agents and ubiquitous sensors. *Procedia Manufacturing*, *52*, 67-72.

Singh, A., Wiktorsson, M., & Hauge, J. B. (2021). Trends In Machine Learning To Solve Problems In Logistics. *Procedia CIRP*, *103*, 67-72.

Giuffrida, N., Fajardo-Calderin, J., Masegosa, A. D., Werner, F., Steudter, M., & Pilla, F. (2022). Optimization and machine learning applied to last-mile logistics: A review. *Sustainability*, *14*(9), 5329.

Rajesh, P., Alam, M., & Tahernezhadi, M. (2020, July). A data science approach to football team player selection. In *2020 IEEE International Conference on Electro Information Technology (EIT)* (pp. 175-183). IEEE.

Andriyanov, N. (2020). Comparative analysis of football statistics data clustering algorithms based on deep learning and Gaussian mixture model. In *CEUR Workshop Proceedings* (Vol. 2667, pp. 71-74).

Zhao, P., & Dong, G. (2022). Analysis of the optimal shooting angle in football matches based on network data mining. *Optik*, *270*, 169925.

Rodrigues, F., & Pinto, Â. (2022). Prediction of football match results with Machine Learning. *Procedia Computer Science*, *204*, 463-470.

Anfilets, S., Bezobrazov, S., Golovko, V., Sachenko, A., Komar, M., Dolny, R., ... & Osolinskyi, O. (2020). Deep multilayer neural network for predicting the winner of football matches. *International Journal of Computing*, *19*(1), 70-77.

Rathi, K., Somani, P., Koul, A. V., & Manu, K. S. (2020). Applications of artificial intelligence in the game of football: The global perspective. *Researchers World*, *11*(2), 18-29.

Majumdar, A., Bakirov, R., Hodges, D., Scott, S., & Rees, T. (2022). Machine Learning for Understanding and Predicting Injuries in Football. *Sports Medicine-Open*, *8*(1), 1-10.

Ali, M., Hassan, M., Kifayat, K., Kim, J. Y., Hakak, S., & Khan, M. K. (2023). Social media content classification and community detection using deep learning and graph analytics. *Technological Forecasting and Social Change*, *188*, 122252.

Arasu, B. S., Seelan, B. J. B., & Thamaraiselvan, N. (2020). A machine learning-based approach to enhancing social media marketing. *Computers & Electrical Engineering*, *86*, 106723.

Li, X., Lu, P., Hu, L., Wang, X., & Lu, L. (2022). A novel self-learning semi-supervised deep learning network to detect fake news on social media. *Multimedia Tools and Applications*, *81*(14), 19341-19349.

Babu, N. V., & Kanaga, E. (2022). Sentiment analysis in social media data for depression detection using artificial intelligence: A review. *SN Computer Science*, *3*(1), 1-20.

Sharma, A., & Shafiq, M. O. (2022). A Comprehensive Artificial Intelligence Based User Intention Assessment Model from Online Reviews and Social Media. *Applied Artificial Intelligence*, 1-26.

Sufi, F., & Alsulami, M. (2022). A Novel Method of Generating Geospatial Intelligence from Social Media Posts of Political Leaders. *Information*, *13*(3), 120.

Majewska, A., & Majewska, A. (2022). Are social media matter for the football club finance?. *Procedia Computer Science*, *207*, 2068-2076.

Plaza, B. (2011). Google Analytics for measuring website performance. *Tourism Management*, *32*(3), 477-481.

Sousa, R., Miranda, R., Moreira, A., Alves, C., Lori, N., & Machado, J. (2021). Software tools for conducting real-time information processing and visualization in industry: An up-to-date review. *Applied Sciences*, *11*(11), 4800.

Nanda, P. (2022). Social Media Analytics for E-Business: Where Are We Heading?. *Organization, Business and Management*, 25.

Sameera, N., & Shashi, M. (2020). Encoding approach for intrusion detection using PCA and KNN classifier. In *Proceedings of the Third International Conference on Computational Intelligence and Informatics* (pp. 187-199). Springer, Singapore.

Barradas, A., Tejeda-Gil, A., & Cantón-Croda, R. M. (2022). Real-Time Big Data Architecture for Processing Cryptocurrency and Social Media Data: A Clustering Approach Based on k-Means. *Algorithms*, *15*(5), 140.