THE PIVOTAL ROLE OF INTERPRETABILITY IN EMPLOYEE ATTRITION PREDICTION AND DECISION-MAKING

Gabriel Marín Díaz, José Javier Galán Hernández

Universidad Complutense de Madrid (Spain)

gmarin03@ucm.es; josejgal@ucm.es

ABSTRACT

This article explores the evolution of machine learning (ML) algorithms, emphasizing the growing importance of interpretability in understanding automated decisions. Progress from early to advanced ML models highlights the need for better performance and adaptability. However, the inherent black-box nature of many ML algorithms raises challenges, underscoring the necessity for interpretability to improve transparency and accountability.

Examining the evolution of interpretability in ML, the article showcases advancements in techniques facilitating human comprehension of decision-making processes. As ML becomes integral across domains, the article underscores the importance of interpretable models to bridge the gap between automated decisions and human understanding.

The article delves into the changing role of humans in decision-making. Despite the efficiency of ML algorithms, the interpretability factor prompts a revaluation of human involvement, necessitating a balanced approach for ethical AI deployment.

Furthermore, the article explores integrating decision-making methods like Analytic Hierarchy Process (AHP) to enhance interpretability. Proposing a framework that combines AHP with interpretable ML models, it suggests a structured approach for human-in-the-loop decision-making while considering feature importance.

KEYWORDS: decision-making, machine learning, XAI, interpretability, AI, AHP.

1. INTRODUCTION

Throughout the historical evolution of Artificial Intelligence (AI) and machine learning (ML) algorithms, the emphasis on interpretability has become increasingly critical, particularly in the dynamic landscape of the business world (Hall, 2022). While the historical narrative traces the roots of interpretability in the context of AI development, its significance in the business realm, and specifically in areas like Human Resources (HR), is a pressing concern (Bandyopadhyay & Jadhav, 2021).

In the contemporary business environment, the adoption of ML algorithms is pervasive, and their applications extend to crucial domains such as HR, where decisions regarding employee management and resource allocation have profound implications. Achieving a balance between predictive power and human-understandable insights becomes paramount, especially when dealing with sensitive areas like employee attrition.

Interpretability is vital in the business context for several reasons. Firstly, businesses need to comply with ethical standards and legal regulations (Bibal et al., 2020). Transparent and interpretable ML models are essential for ensuring that decisions related to hiring, promotions, and terminations align with fairness and non-discrimination principles. Secondly, in HR, the ability to explain why a particular decision was made becomes crucial for building trust among employees and stakeholders (Mishra, 2013). For instance, if an algorithm predicts an employee is likely to leave the company, it is imperative to understand the features contributing to this prediction to take appropriate actions.

Consider a scenario where a company utilizes an ML algorithm to predict employee attrition. An interpretable model not only provides accurate predictions but also offers explanations for those predictions. This transparency allows HR professionals to understand the factors influencing an employee's likelihood of leaving, enabling them to intervene proactively. Interpretability, in this context, becomes a tool for strategic workforce planning, talent retention, and fostering a more inclusive workplace culture (Marín Díaz et al., 2023).

In HR decision-making, interpretability aids in justifying and fine-tuning models based on real-world observations, aligning them with organizational values (Srivastava & Eachempati, 2021). It empowers HR professionals to leverage the strengths of ML models while retaining human oversight in critical decision-making processes. The interpretability of algorithms in HR ensures that the human touch remains integral, fostering a collaborative and ethical approach to workforce management.

Interpretability is indispensable in the business landscape, particularly in critical areas like HR, where algorithmic decisions impact the livelihoods and well-being of employees. By shedding light on the decision-making process, interpretable ML models not only enhance trust but also contribute to strategic and ethical human resource management.

This paper addresses the challenges associated with interpretability in machine learning (ML) models, following a structured framework. Section 2 reviews the current state of eXplainable Artificial Intelligence (XAI) in the business domain. In Section 3, a methodology is proposed, focusing on handling algorithmic explainability while incorporating a crucial aspect—human decision-making. Section 4 presents a real-world use case illustrating the application of the proposed methodology. Finally, Section 5 provides conclusions and outlines future directions for research and development.

2. RELATED WORK

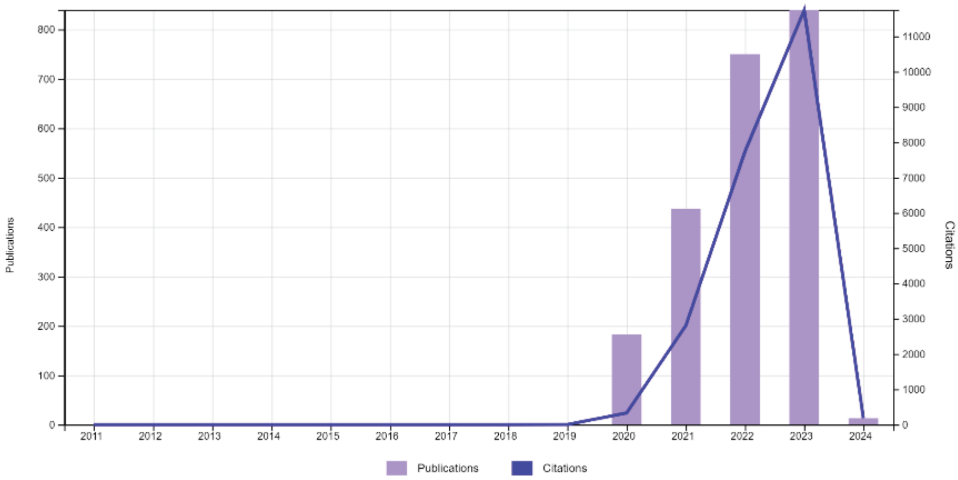
Employee attrition has emerged as a focal point for organizations, especially in the ever-evolving landscape of contemporary workplaces. Examining this from a psychological standpoint, numerous influential factors contribute to an individual's decision to depart from their current employment, especially in the technology sector (Colomo-Palacios et al., 2014).

Within this rapidly changing industry, newly hired employees often prioritize elements such as job satisfaction, a conducive work environment, and substantial financial compensation over traditional notions of stability and long-term commitment. Attrition is influenced not only by intrinsic aspects of the job role but also by organizational culture, growth prospects, and the alignment of individual values with the company's mission (Climek et al., 2022).

Comprehending the intricate dynamics and psychological foundations of employee attrition, particularly in technology-driven sectors, is imperative for crafting effective retention strategies.

The growing attention towards eXplainable Artificial Intelligence (XAI) highlights the increasing emphasis on creating AI systems that are understandable, particularly in situations where decisions have broad implications for individuals or society. Striking the appropriate equilibrium between predictive accuracy and interpretability remains an enduring challenge in the field. This equilibrium holds significant importance in building trust and gaining acceptance for AI systems in practical applications, ranging from healthcare to finance and beyond. Figure 1 visually depicts the evolution in the volume of studies addressing the interpretability of algorithms over time, TS = ("eXplainable Artificial Intelligence " or "XAI").

Figure 1. Studies addressing the interpretability of algorithms (2,221 publications).



Source: self-elaboration based on Web of Science (2024)

As observed in Table 1, the number of publications is centered around scientific areas, although the practical implementation of eXplainable Artificial Intelligence (XAI) models applied to the business world is not significant.

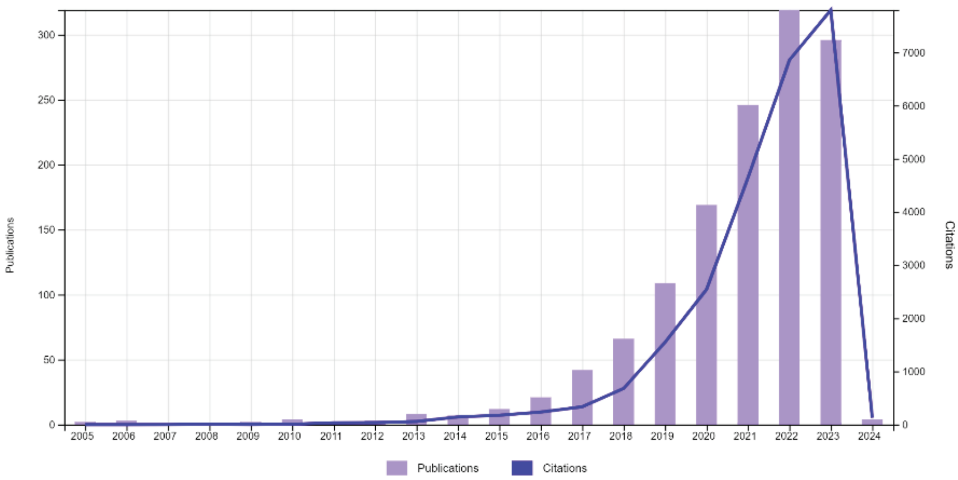
Table 5. Publications by research areas.

Area	Publications	%
Computer Science	1,646	74.111
Engineering	772	34.759
Mathematics	571	25.709
Mathematical Computational Biology	516	23.233
Communication	300	13.507

Source: self-elaboration based on Web of Science (2024)

Next, we proceed to analyse publications related to Machine Learning and Artificial Intelligence applied to the Human Resources sector, Figure 2, without currently delving into studies on interpretability in this area, TS = ("Machine Learning" or "Artificial Intelligence") AND TS = ("Resource Humans" or "HR").

Figure 2. Studies addressing the ML / AI applied to the Human Resources (1,314 publications).



Source: self-elaboration based on Web of Science (2024)

Finally, we focus on studies related to eXplainable Artificial Intelligence (XAI) that directly impact the Human Resources field, Table 2, TS = ("eXplainable Artificial Intelligence " or "XAI") AND TS = ("Resource Humans" or "HR").

Table 2. Publications XAI and Human Resources.

Publications
Applying XAI to an AI-based system for candidate management to mitigate bias and discrimination in hiring (Hofeditz et al., 2022)
Analyzing Employee Attrition Using Explainable AI for Strategic HR Decision-Making (Marín Díaz et al., 2023)

Source: self-elaboration based on Web of Science (2024)

As evidenced by Figure 2 and Table 2, the utilization of predictive models in the Human Resources domain is extensively documented. However, it is noteworthy that articles specifically addressing interpretability amount to a total of 2.

3. METHODOLOGY

3.1. Interpretable Machine Learning

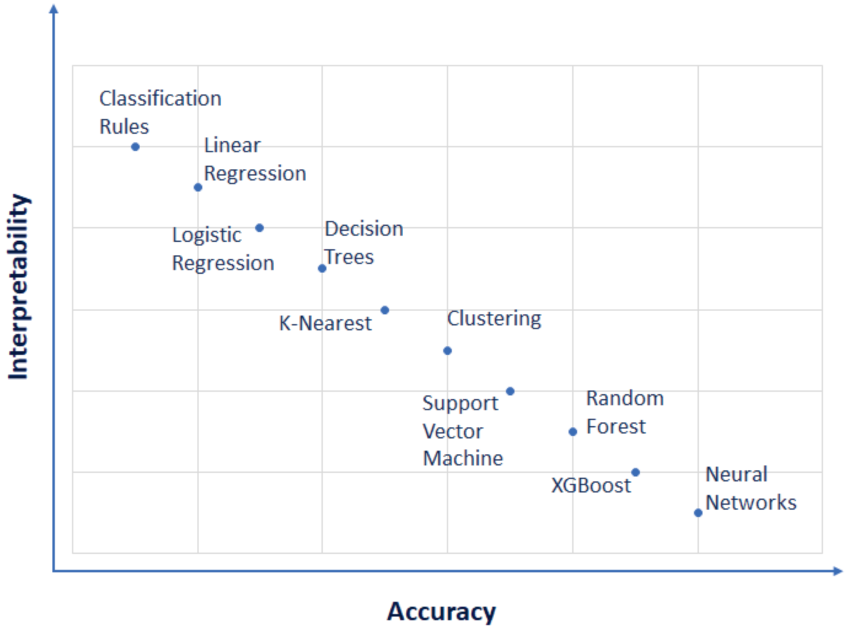
Interpretable Machine Learning (IML) is a pivotal component when decisions carry significant implications. This in-depth exploration will delve into various methods of interpretability and their practical implementation in environments where decision-making is of paramount importance.

When examining the interpretability of machine learning models, scholars commonly classify them into two fundamental categories (Carvalho et al., 2019). Firstly, 'transparent models,' often referred to as 'white box models,' aim to establish a clear connection between input variables and resulting outputs. On the other hand, 'opaque models,' or 'black box models,' lack easily interpretable decision rules. It is noteworthy that even within the realm of transparent models, interpretability remains a topic of ongoing discussion, as highlighted in a distinct study that raises questions about their interpretability (Lipton, 2018).

Figure 3 portrays an inverse correlation between interpretability and accuracy, emphasizing the intricate challenge of striking a balance between model interpretability and predictive precision (Molnar, 2019). Distinguishing between these model types provides valuable insights into the spectrum of interpretability within the domain of machine learning. This nuanced perspective enhances our comprehension of model behaviours and performance, contributing depth to the ongoing discourse on interpretability.

It is essential to acknowledge that the presence of bias and noise in data can distort interpretations. Addressing these issues is imperative before embarking on any interpretability analysis. Data cleaning techniques, such as class balancing, play a crucial role in mitigating bias, while noise removal ensures that interpretations rely on reliable information (Gilpin et al., 2019). The following outlines various methods for interpreting algorithms.

Figure 3. Interpretability and Accuracy.



Source: self-elaboration based on (Duval, 2019).

Feature Importance Analysis: Quantifies the impact of individual features on model predictions, often expressed through coefficients or weights (Perisic & Pahor, 2020). Commonly applied in linear models, providing explicit feature contributions.

Decision Tree Structure Analysis: Investigates the hierarchical decision-making process of decision trees, offering insight into feature importance and splits. Decision trees and ensemble methods like Random Forest (Freitas, 2014).

Gradient-Based Attribution Methods: Leverages partial derivatives to attribute model predictions to specific features, enhancing understanding of feature contributions. Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) (Molnar, 2019).

Residual Analysis: Examines the discrepancies between predicted and actual outcomes, aiding in the identification of systematic errors (Sangeetha & Prasad, 2006). Residual plots help reveal patterns and model shortcomings.

Visualization of Feature Relationships: Utilizes graphical representations to depict the relationships between features and model predictions. Scatter plots and heatmaps for intuitive interpretation (Goldstein et al., 2015).

Permutation Feature Importance Analysis: Evaluates the importance of features by systematically permuting their values and measuring the impact on model performance (Altmann et al., 2010). A rigorous approach to discerning feature importance under various perturbations.

Variable Profiling and Sensitivity Analysis: Explores how model predictions evolve as individual features undergo controlled variations (Montavon et al., 2018). Comprehensive sensitivity analyses, assessing global and local model responses.

Analysis of Intermediate Model Representations: Investigates the transformations and representations within intermediate layers of complex models, shedding light on information processing (Goodrich, 2010). In-depth examination of neural network architectures.

Association Rule Mining: Identifies frequent patterns and associations in the data, contributing to the understanding of feature interactions (Hsieh, 2004). Application of the Apriori algorithm to discover significant rule sets.

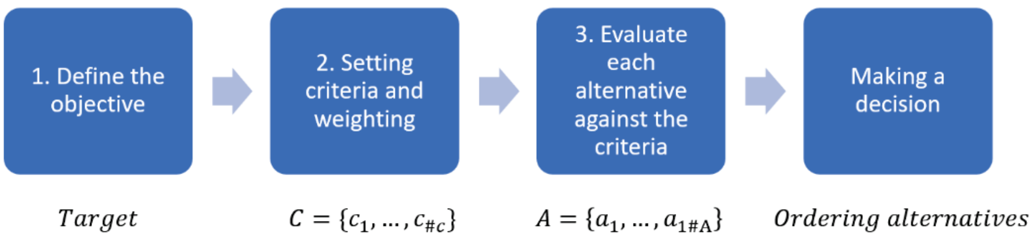
While interpretable machine learning enhances decision-making, challenges must be navigated. Striking a balance between accuracy and interpretability is an ongoing challenge, and understanding the trade-offs is crucial. Additionally, the ethical considerations of interpretable models, especially in sensitive decision-making contexts, require careful attention.

3.2. Analytic Hierarchy Process (AHP)

The Analytic Hierarchy Process (AHP), introduced by Thomas Saaty (Thomas L. Saaty, 2008), is a decision-making technique designed to address complex and multi-criteria problems. This method involves hierarchically structuring decision factors and systematically comparing available options, applicable across diverse domains such as business, engineering, healthcare, and environmental planning.

Considerations for applying AHP include the number of experts involved in decision-making, varying from a single expert to a group. Engaging multiple experts incorporates diverse perspectives, enhancing the robustness of decisions. The decision environment, classified into structured and unstructured, influences AHP's effectiveness. In structured environments, well-defined and quantifiable criteria facilitate systematic comparison, while unstructured environments require expert judgment and qualitative assessments, Figure 4.

Figure 4. Analytic Hierarchy Process (AHP).



Source: self-elaboration based on (Saaty, 1980).

AHP offers a versatile decision-making approach adaptable to various scenarios, specifically tailored for intricate decision scenarios with multiple criteria (Cid-López et al., 2016). Factors like multiple experts, decision environment, and the number of criteria are crucial for effective AHP utilization, allowing decision-makers to categorize problems and apply appropriate techniques.

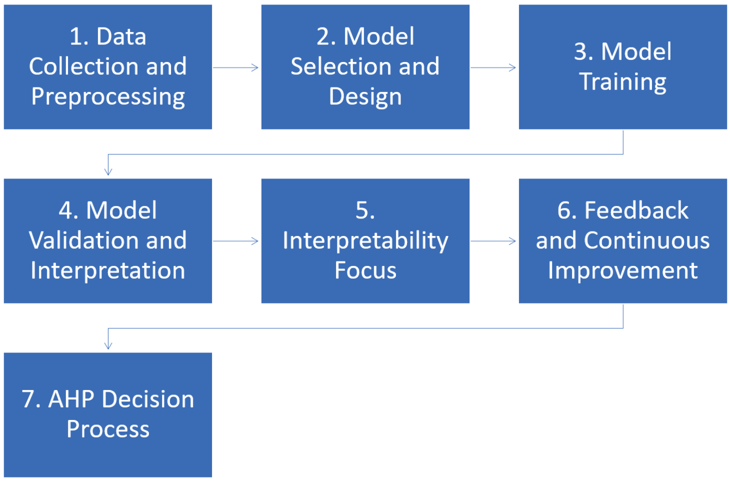
3.3. Proposed Model

Through the AHP model, we can derive the various weights of the most significant features obtained via XAI and incorporate the human factor into the decision-making process. The decision regarding the hiring of a specific profile is not solely determined by AI. In our case, we conduct an analysis of the model's most crucial characteristics related to employee attrition. These features play a pivotal role in the decision-making process for hiring new employees, emphasizing those essential characteristics to prevent attrition.

In this decision-making process, guided by the insights provided by the AI model, we can apply the Analytic Hierarchy Process (AHP). This structured approach allows us to prioritize and weigh the identified characteristics, aligning with the model's recommendations to enhance the hiring decision-making process.

The process is detailed below and is depicted in Figure 5.

Figure 5. Proposed Model (AI + XAI + AHP).



Source: self-elaboration based on (Shafique & Qaiser, 2014).

Data Collection and Preprocessing: Begin with obtaining relevant data and preparing it for analysis. During preprocessing, address issues such as outliers, missing data, and normalization.

Model Selection and Design: Choose an appropriate model for the problem and design its architecture. Consider interpretable models, such as decision trees or linear regressions, when possible.

Model Training: Use labeled data to adjust the model's parameters. Validate and fine-tune performance on a validation set.

Model Validation and Interpretation: Evaluate the model on an independent test set to measure its performance. Interpret how the model makes decisions in specific cases.

Model Deployment: Implement the model in an operational environment for real-time predictions.

Continuous Monitoring: Monitor the model's performance in production and adjust as needed.

Feedback and Continuous Improvement: Gather user feedback and adjust the model based on new needs or changes in the data.

AHP Decision Process: Once the most relevant features of the predictive model are identified, apply these features in a decision-making process using the Analytic Hierarchy Process (AHP).

By integrating AHP in the final stages of the process, after selecting the most relevant characteristics from the predictive model, we can enhance the decision-making process with a systematic and interpretable approach.

4. PRACTICAL APPLICATION

The data were gathered from the publicly available IBM HR database (Kaggle HR Analytic Data Set, n.d.), and Table 3 enumerates the features comprising the dataset.

Table 3. Data Set IBM HR.

Features	
Age	Monthly Income
Attrition	Monthly Rate
Business Travel	Number of Companies Worked
Daily Rate	Over18
Department	Over Time
Distance from Home	Percent Salary Hike
Education	Performance Rating
Education Field	Relationship Satisfaction
Employee Count	Standard Hours
Employee Number	Stock Option Level
Environment Satisfaction	Total Working Years
Gender	Training Times Last Year
Hourly Rate	Work Life Balance
Job Involvement	Years at Company
Job Level	Years in Current Role
Job Role	Years since Last Promotion
Job Satisfaction	Years with Current Manager
Marital Status	

Source: IBM HR (*Kaggle HR Analytic Data Set*, n.d.)

After completing the research process, exploratory analysis, and predictive modeling, the model that best fits our observations is XGBoost, with an Accuracy Mean of 85.91. The model is a black-box, and therefore, we apply algorithm interpretability to understand the most relevant features influencing employee turnover.

Figure 6. Features Importance, ELI5.

Weight	Feature
0.0867 ± 0.0119	OverTime
0.0373 ± 0.0113	MonthlyIncome
0.0089 ± 0.0072	DailyRate
0.0056 ± 0.0033	DistanceFromHome
0.0052 ± 0.0010	RelationshipSatisfaction
0.0051 ± 0.0015	JobSatisfaction
0.0047 ± 0.0019	NumCompaniesWorked
0.0043 ± 0.0029	MonthlyRate
0.0041 ± 0.0015	MaritalStatus_Single
0.0041 ± 0.0026	StockOptionLevel
0.0033 ± 0.0020	Age
0.0025 ± 0.0016	EnvironmentSatisfaction
0.0021 ± 0.0008	PercentSalaryHike
0.0017 ± 0.0008	YearsAtCompany
0.0014 ± 0.0010	JobInvolvement
0.0014 ± 0.0016	YearsSinceLastPromotion
0.0012 ± 0.0008	BusinessTravel_Travel_Frequently
0.0010 ± 0.0000	EducationField_Technical Degree
0.0010 ± 0.0017	HourlyRate
0.0008 ± 0.0008	JobRole_Research Scientist
	… 24 more …

After completing the research process, exploratory analysis, and predictive modeling, the model that best fits our observations is XGBoost, with an Accuracy Mean of 85.91. The model is a black-box, and therefore, we apply algorithm interpretability to understand the most relevant features influencing employee turnover.

Once the importance of features is visualized through ELI5 (Molnar, 2019), it can be observed that the determining variables in a selection process would be related to income level. We need to offer a contract that is economically competitive. Age shows an inverse correlation with turnover – the higher the age, the lower the probability of abandonment. The same holds true for marital status; singles exhibit a higher propensity for turnover. Avoiding overloading workers with extra hours is crucial; we seek a positive attitude and satisfaction with the work environment. Additionally, minimizing the commuting distance from home is essential.

This same process can help us assess potential departures among our personnel. We must take care of workers facing work overload or earning below-average incomes, especially those with longer-than-average commuting distances from home. As observed, interpretability provides a powerful mechanism for determining the most crucial criteria in a selection process or in guiding and supporting workers.

The AHP model, once the criteria for personnel selection are defined, allows us to obtain a criterion weighting according to values provided by the predictive and interpretable model. This facilitates the decision-making process, enabling the selection of the best alternative for the vacant position.

For the case study, we conducted the AHP analysis to determine a selection process with three potential candidates, considering the following criteria and alternatives, as well as the following pair-wise comparison matrices.

Figure 7. AHP Analysis, selection process.

Criteria	
C1	Age
C2	Level of Specialization
C3	Distance from Home
C4	Years of Work
C5	Economic Aspiration

	Alternatives				
	C1	C2	C3	C4	C5
A1	30	4	10	10	60.000,00 €
A2	35	3	30	5	65.000,00 €
A3	42	3	12	7	55.000,00 €

	Criteria weighting					
	C1	C2	C3	C4	C5	Average Vector
C1	1,00	0,20	3,00	0,33	0,33	9,76%
C2	5,00	1,00	5,00	3,00	3,00	44,28%
C3	0,33	0,20	1,00	0,33	0,33	6,01%
C4	3,00	0,33	3,00	1,00	3,00	23,92%
C5	3,00	0,33	3,00	0,33	1,00	16,03%

	C1 = Age			
	T1	T2	T3	Average Vector
T1	1,00	0,33	0,20	10,62%
T2	3,00	1,00	0,33	26,05%
T3	5,00	3,00	1,00	63,33%

	C4 = Years of Work			
	T1	T2	T3	Average Vector
T1	1,00	5,00	3,00	63,33%
T2	0,20	1,00	0,33	10,62%
T3	0,33	3,00	1,00	26,05%

	C2 = Level of Specialization			
	T1	T2	T3	Average Vector
T1	1,00	3,00	3,00	60,00%
T2	0,33	1,00	1,00	20,00%
T3	0,33	1,00	1,00	20,00%

	C5 = Economic Aspiration			
	T1	T2	T3	Average Vector
T1	1,00	3,00	0,33	26,05%
T2	0,33	1,00	0,20	10,62%
T3	3,00	5,00	1,00	63,33%

	C3 = Distance from Home			
	T1	T2	T3	Average Vector
T1	1,00	5,00	1,00	47,96%
T2	0,20	1,00	0,33	11,50%
T3	1,00	3,00	1,00	40,55%

After completing the entire process, alternative 1 is considered the optimal choice with a weight of 49.81%, the second option corresponds to alternative 3, with a weight of 33.86%, and finally, alternative 2 with a weight of 16.33%. Therefore, upon concluding the process, it can be asserted that, considering the interpretability applied to employee turnover, along with the AHP method adhering to the recommended criteria for the hiring process, we opt for the most suitable candidate.

5. CONCLUSIONS

In this study, eXplainable Artificial Intelligence (XAI), was utilized to address the issue of employee turnover. The use of interpretable techniques allowed for the identification and measurement of the importance of various characteristics related to this phenomenon.

Through the measurement of feature importance with ELI5, a detailed investigation was conducted on the criteria that could trigger employee turnover. This approach provided a solid foundation for developing a more informed personnel selection process aimed at preventing employee attrition.

The application of interpretability in Machine Learning (ML) algorithms emerges as a crucial component in decision-making. The ability to understand and explain model decisions not only enhances confidence in these models but also facilitates the adoption of more ethical and well-founded decisions.

This work underscores the imperative need to consider interpretability in ML algorithms, not only for its practical utility but also for the associated ethical implications. Opacity in automated decisions can lead to unexpected consequences and a lack of accountability. The adoption of interpretable approaches aligns with the pursuit of transparency and responsibility in algorithm implementation.

The practical application of interpretability in the context of employee attrition highlights its utility in identifying and understanding determining factors. This knowledge is valuable not only for decision-making in human resources but also contributes to talent retention and strengthens organizational policies.

The integration of Explainable Artificial Intelligence (XAI) and the application of Decision Theory, specifically the Analytic Hierarchy Process (AHP), bestow decisions with an enriching collaboration between humans and machines. This synergistic approach not only enhances interpretability in complex decision-making scenarios but also underscores the symbiotic relationship between human insight and machine-driven analyses, contributing to a more informed and ethical decision landscape.

REFERENCES

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: A corrected feature importance measure. *Bioinformatics*, *26*(10), 1340–1347. https://doi.org/10.1093/bioinformatics/btq134

Bandyopadhyay, N., & Jadhav, A. (2021). Churn Prediction of Employees Using Machine Learning Techniques. *Tehnicki Glasnik*, *15*(1), 51–59. https://doi.org/10.31803/tg-20210204181812

Bibal, A., Lognoul, M., de Streel, A., & Frénay, B. (2020). Legal requirements on explainability in machine learning. *Artificial Intelligence and Law*, *0123456789*. https://doi.org/10.1007/s10506-020-09270-4

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics (Switzerland)*, *8*(8), 1–34. https://doi.org/10.3390/electronics8080832

Cid-López, A., Hornos, M. J., Carrasco, R. A., & Herrera-Viedma, E. (2016). Applying a linguistic multi-criteria decision-making model to the analysis of ICT suppliers' offers. *Expert Systems with Applications*, *57*, 127–138. https://doi.org/10.1016/j.eswa.2016.03.025

Climek, M., Henry, R., & Jeong, S. (2022). Integrative literature review on employee turnover antecedents across different generations: commonalities and uniqueness. *European Journal of Training and Development*, *ahead-of-p*(ahead-of-print). https://doi.org/10.1108/EJTD-05-2021-0058

Colomo-Palacios, R., Casado-Lumbreras, C., Misra, S., & Soto-Acosta, P. (2014). Career Abandonment Intentions among Software Workers. *HUMAN FACTORS AND ERGONOMICS IN MANUFACTURING & SERVICE INDUSTRIES*, *24*(6), 641–655. https://doi.org/10.1002/hfm.20509

Duval, A. (2019). *Explainable Artificial Intelligence (XAI) Explainable Artificial*. *April*. https://doi.org/10.13140/RG.2.2.24722.09929

Freitas, A. A. (2014). Comprehensible classification models: a position paper. *ACM SIGKDD Explorations Newsletter*, *15*(1), 1–10.

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2019). Explaining explanations: An overview of interpretability of machine learning. *Proceedings - 2018 IEEE 5th International Conference on Data Science and Advanced Analytics, DSAA 2018*, 80–89. https://doi.org/10.1109/DSAA.2018.00018

Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. *Journal of Computational and Graphical Statistics*, *24*(1), 44–65. https://doi.org/10.1080/10618600.2014.907095

Goodrich, M. T. (2010). Data Structures and Algorithms in Python. *Wiley*, *53*(9), 1689–1699. http://arxiv.org/abs/1011.1669%0Ahttp://dx.doi.org/10.1088/1751-8113/44/8/085201

Hall, P. (2022). *Machine Learning for High-Risk Applications*.

Hofeditz, L., Clausen, S., Rieß, A., Mirbabaie, M., & Stieglitz, S. (2022). Applying XAI to an AI-based system for candidate management to mitigate bias and discrimination in hiring. *Electronic Markets*, *32*(4), 2207–2233. https://doi.org/10.1007/s12525-022-00600-9

Hsieh, N. C. (2004). An integrated data mining and behavioral scoring model for analyzing bank customers. *Expert Systems with Applications*, *27*(4), 623–633. https://doi.org/10.1016/j.eswa.2004.06.007

Kaggle HR Analytic Data Set. (n.d.). https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset

Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM*, *61*(10), 35–43. https://doi.org/10.1145/3233231

Marín Díaz, G., Galán Hernández, J. J., & Galdón Salvador, J. L. (2023). Analyzing Employee Attrition Using Explainable AI for Strategic HR Decision-Making. *Mathematics*, *11*(22). http://doi.org/10.3390/math11224677

Mishra, D. (2013). Review of literature on factors influencing attrition and retention. *International Journal of Organizational Behaviour & Management Perspectives*, *2*(3), 435–445.

Molnar, C. (2019). Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. *Book*, 247. https://christophm.github.io/interpretable-ml-book

Montavon, G., Samek, W., & Müller, K. R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing: A Review Journal*, *73*, 1–15. https://doi.org/10.1016/j.dsp.2017.10.011

Perisic, A., & Pahor, M. (2020). Extended RFM logit model for churn prediction in the mobile gaming market. *Croatian Operational Research Review*, *11*(2), 249–261. https://doi.org/10.17535/crorr.2020.0020

Saaty, T. L. (1980). The analytic hierarchy process : planning, priority setting, resource allocation LK - https://ucm.on.worldcat.org/oclc/911278091. In *TA - TT* -. McGraw-Hill International Book Co.

Sangeetha, V., & Prasad, K. J. R. (2006). Deep residual learning for image recognition. *Indian Journal of Chemistry - Section B Organic and Medicinal Chemistry*, *45*(8), 1951–1954. https://doi.org/10.1002/chin.200650130

Shafique, U., & Qaiser, H. (2014). A Comparative Study of Data Mining Process Models (KDD , CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, *12*(1), 217–222. http://www.ijisr.issr-journals.org/

Srivastava, P. R., & Eachempati, P. (2021). Intelligent Employee Retention System for Attrition Rate Analysis and Churn Prediction: An Ensemble Machine Learning and Multi- Criteria Decision-Making Approach. *Journal of Global Information Management*, *29*(6), 1–29. https://doi.org/10.4018/JGIM.20211101.oa23

Thomas L. Saaty. (2008). Decision making with the analytic hierarchy process. *Journal of Manufacturing Technology Management*, *26*(6), 791–806. https://doi.org/10.1108/JMTM-03-2014-0020