

Article

Not peer-reviewed version

Integrating Explainable AI (XAI) and NCA-Validated Clustering for an Interpretable Multi-layered Recruitment Model

[Marcin Nowak](#)^{*} and [Marta Pawłowska-Nowak](#)

Posted Date: 17 December 2025

doi: 10.20944/preprints202512.1497.v1

Keywords: machine learning; Explainable Artificial Intelligence; human resources management; ethics in HRM; recruitment



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Integrating Explainable AI (XAI) and NCA-Validated Clustering for an Interpretable Multi-Layered Recruitment Model

Marcin Nowak * and Marta Pawłowska-Nowak

Poznan University of Technology, Faculty of Engineering Management

* Correspondence: marcin.nowak@put.poznan.pl

Abstract

This article proposes an interpretable, multi-layered recruitment model that balances predictive performance with decision transparency in AI-supported HR processes, addressing risks related to opacity, auditability, and ethically sensitive decision-making. The architecture combines an expert rule layer for minimum-threshold screening, an unsupervised clustering layer to structure candidate profiles and generate pseudo-labels, and a supervised classification layer trained and evaluated via repeated k-fold cross-validation. Model behavior is explained using SHAP to identify feature contributions to cluster assignment, and cluster quality is additionally diagnosed using Necessary Condition Analysis (NCA) to assess minimum competency requirements for attaining a target overall quality level. The approach is illustrated in a Data Scientist recruitment case study, where centroid-based clustering predominates (K-Means is most frequently selected), while linear classifiers show the highest effectiveness and stability (logistic regression performs best). SHAP highlights competencies that differentiate candidates beyond the initial threshold, and NCA further distinguishes candidates within the recommended cluster by identifying profiles that meet (or fail) the necessary-condition bottleneck. The proposed framework is replicable and supports transparent, auditable recruitment decisions.

Keywords: machine learning; Explainable Artificial Intelligence; human resources management; ethics in HRM; recruitment

1. Introduction

The rapid development of artificial intelligence has made machine learning-based tools an integral part of human resource management processes-particularly recruitment. Organizations are increasingly using algorithms for automatic application filtering, candidate profiling, and predicting job fit, which helps reduce costs and shorten time-to-hire. At the same time, risks associated with model opacity, the potential reinforcement of biases encoded in historical data, and the need to align AI systems with regulatory requirements-such as the EU's AI Act-are becoming more pronounced. In this context, solutions that combine high predictive performance with the ability to explain and audit algorithmic decisions gain crucial importance. Despite the growing body of research on AI applications in HR, there remains a lack of comprehensive, formally described recruitment architectures that integrate expert knowledge, unsupervised data structuring methods, supervised classification models, and Explainable AI techniques, while also subjecting the results to additional validation through Necessary Condition Analysis (NCA). The aim of this article is to address this gap by proposing a multi-layered, interpretable recruitment model based on XAI methods and clustering validated using NCA. The following sections provide a literature review on AI applications in HR and related ethical challenges, present the formal structure of the proposed model, and illustrate its functioning through a case study on recruitment for a Data Scientist position. The article concludes

with a summary discussing the theoretical contribution, practical implications, limitations of the approach, and directions for future research.

2. Literature Review

Artificial intelligence (AI) plays an increasingly significant role in human resource management (HRM), transforming the processes of recruitment, selection, and employee development [1]. As early as the 1980s, the first computer-based tools supporting recruitment emerged-rule-based applicant tracking systems (ATS) designed to manage the influx of candidate resumes [2]. In the following decades, technological progress accelerated: in the 1990s, the digitization of HR data and the advent of the internet enabled the creation of candidate databases, while the early 2000s saw the emergence of professional networking platforms (e.g., LinkedIn) and more advanced ATS capable of filtering applications based on keywords. Simultaneously, the first recruitment chatbots were introduced, automating communication with candidates. However, concerns were already being raised at that time-algorithms trained on historical data could reinforce existing biases and unfairness in the recruitment process [3]. Since around 2015, there has been a sharp increase in the use of machine learning and deep learning in HR [4,5]. AI has evolved from a tool for administrative support into a strategic component of personnel decision-making.

Recruitment today employs intelligent talent search systems that scan the internet and CV databases for suitable candidates, initial screening using AI (e.g., job interviews conducted by chatbots), and even controversial analyses of speech, facial expressions, and behavior in video interview recordings [6]. In the areas of employee development and retention, predictive algorithms are used, such as models forecasting employee attrition risk (supporting retention efforts) and systems recommending career paths and training tailored to an employee's profile [7]. AI also supports the monitoring of compliance and procedural adherence in HR. These tools provide HR departments with vast amounts of data and insights, enabling more informed, data-driven personnel decisions. Consequently, the HR function has evolved from a purely administrative role to a strategic partner supporting business goals [8].

Surveys indicate a growing adoption of AI-globally, nearly half of organizations report already using AI tools in HR functions [9], and most of the remainder plan to implement them in the near future. AI enables companies to automate tedious administrative tasks, accelerate candidate screening (algorithms can filter hundreds of resumes in a fraction of the time it takes a human), and reduce recruitment costs [9]. For example, it is estimated that the use of AI in recruitment can lower hiring costs by up to ~30% and shorten time-to-hire by 50% [9]. As a result, AI has become a key driver of HR transformation, ushering in the era of data-driven recruitment [10].

Despite its undeniable benefits, the use of AI in HRM raises significant ethical challenges. The most serious of these is the risk of bias and discrimination. Algorithms trained on historical HR data can unknowingly adopt and reinforce the biases present in those data-if, for instance, a certain group of candidates was favored in the past, the system may begin to replicate these preferences automatically [11,12]. A well-known example is Amazon's 2018 case, where an internal AI system for resume evaluation learned to reject female candidates. The model, trained on resumes historically submitted to the company (which were predominantly from men), identified female gender as a factor lowering candidate quality-penalizing occurrences of the word "women's" (e.g., "women's chess club") in resumes and downgrading graduates of women's colleges [13]. Despite efforts to correct the algorithm, it found alternative ways to replicate the bias, ultimately leading to the project's abandonment. This case highlights how an opaque AI model can discriminate against a specific group and how difficult it is to detect and eliminate such behavior without appropriate explainability mechanisms. Unfortunately, gender is only one of many potential sources of bias-others include age, ethnicity, and disability-which an uncontrolled AI may begin to treat as selection criteria if they correlate with success in historical recruitment data.

Accountability and human oversight of decisions is a crucial ethical aspect of using AI in HR. Critics argue that key decisions affecting people's lives-such as hiring or rejecting a candidate-should

not be left solely to machines, particularly since algorithms lack human empathy and situational context [14]. Philosophers in the field of technology ethics contend that AI is incapable of reflecting the social values and intuition necessary in evaluating employees [14]. There have been instances where systems recommended unsuitable candidates due to errors or limitations in training data [15]. For this reason, the principle of “human-in-the-loop” is being introduced-ensuring human supervision over critical decisions and the ability to verify and correct AI recommendations [16].

The European Union has developed the Artificial Intelligence Act (AI Act), which classifies AI systems used in employment as high-risk. This will entail, among other things, requirements for transparency, documentation, impact assessments regarding individuals’ rights, and human oversight of final decisions [17]. Non-compliance with these regulations may result in severe penalties-in the EU, fines of up to 6–7% of a company’s global turnover are planned [18]. This regulatory trend clearly indicates that the social acceptance of AI in HR will depend on meeting ethical requirements-algorithms must operate transparently, fairly, and with respect for candidates’ rights, or their use may be legally restricted.

In light of the above challenges, the interpretability of AI models used in HRM becomes crucial. Interpretability refers to a model’s ability to present its functioning in a form understandable to humans-allowing one to trace why a particular decision was made [19]. This concept is closely tied to the field of XAI (Explainable Artificial Intelligence), a set of methods that make it possible to explain the behavior of even complex, black-box algorithms. The goal of XAI is to provide users with clear and detailed justifications for AI-driven decisions [20]. For example, in a recruitment context, candidates rejected by an AI system may want to know which specific aspects of their resume led to the negative outcome-so they can assess whether the process was fair and understand what to improve in future applications. Likewise, HR managers relying on AI recommendations should be able to understand on what basis the algorithm deemed a candidate suitable for a position-this increases trust in the system and helps detect potential errors or biases.

The application of XAI in HR practice brings tangible benefits. First, it increases trust in AI systems. When a recruiter or manager receives a recommendation from an algorithm (e.g., a top 5 list of candidates for a position), the ability to view the justification-such as candidate A being highlighted due to strong alignment with competencies X and Y, and candidate B having excellent experience in industry Z-makes it easier to trust the recommendation or consciously challenge it. Research has shown that explanations of AI decisions improve acceptance and understanding among business users [21]. Second, transparency enhances candidates’ perception of fairness in the process. When applicants have access to justifications (e.g., in the form of feedback: “Your application was rejected because you lack required experience in technology X”), they perceive the recruitment process as more just and are more likely to accept the outcome-even if it is negative for them. Third, XAI enables the identification and correction of biases. By analyzing explanations across many decisions, one can detect systematic undesirable patterns-for example, if a model consistently rates candidates from a particular group lower, the explanations will reveal which features lead to this outcome. This allows engineers and HR specialists to intervene (e.g., by removing or adjusting the influence of a given feature, or retraining the dataset) to improve the model’s fairness. In this way, interpretability becomes a tool for ensuring ethical AI behavior-producing models that are more resistant to bias and aligned with the principle of equal treatment of candidates.

A review of the literature reveals that research on the application of AI in HRM tends to focus primarily on two relatively separate areas: first, on developing increasingly effective predictive models to support recruitment decisions, and second, on identifying ethical risks, biases, and general calls for algorithmic transparency. While studies on XAI in HR are emerging, along with isolated case-based examples illustrating the functioning of specific explainability techniques, these typically address a single layer of the model (e.g., classifier explanation) and are not embedded in a coherent, multi-module recruitment architecture. At the same time, the approach based on Necessary Condition Analysis (NCA) remains relatively rare in HRM literature and is mostly used in strictly theoretical research rather than as an integral part of validating algorithmic selection systems.

There is therefore a lack of models that combine: (1) an exploratory layer for unsupervised candidate structuring, (2) an operational layer of supervised classification, (3) a feature-level explainability layer, and (4) a quality validation layer from the perspective of necessary conditions-while also being formally described and replicable in recruitment practice. This article addresses that gap by proposing a multi-layered, fully formalized recruitment model that integrates XAI methods (SHAP) with clustering and NCA. The model ensures both high predictive accuracy and transparent, auditable, and regulation-compliant interpretation of candidate-related decisions.

3. Materials and Methods

The complexity of recruitment processes-which involve multidimensional relationships between competencies, experience, and cultural fit-requires modeling approaches similar to those used in socio-economic systems characterized by high levels of uncertainty [22]. In this context, the proposed recruitment model is based on a multi-layered analytical architecture integrating expert rules, unsupervised learning methods, supervised classification models, and Explainable AI (XAI) techniques. The recruitment model is designed to simultaneously ensure high predictive performance and full interpretability of decisions. The developed model can be represented as a procedure consisting of five steps.

Step 1. Developing a Set of Evaluation Criteria for Candidate Application Forms

In this step, the recruiter must define the criteria for evaluating candidates for a specific job position. The nature of these criteria is flexible in the sense that candidate responses may take the form of both quantitative and qualitative variables, and may be expressed on either continuous or discrete scales.

Step 2. Defining Minimum Thresholds for Evaluation Criteria Enabling Interview Invitation (Rule-Based Layer – Preliminary Filtering)

Step 2 reflects the expert requirements of the recruiter in the form of a set of deterministic rules. Each candidate is evaluated against the minimum criteria necessary to perform the given professional role. Candidates who do not meet even one mandatory condition are automatically assigned to the rejected class (red cluster). The rule layer can be described by the following mathematical structure:

Let (1):

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{id}) \in \mathbb{R}^d \quad (1)$$

denote the feature vector of candidate i , and let

$$\mathcal{R} = \{r_1, r_2, \dots, r_k\} \quad (2)$$

be the set of necessary rules defined by the domain expert (2).

Each rule takes the form (3):

$$r_j(\mathbf{x}_i) = \begin{cases} 1 & \text{if condition } j \text{ is satisfied,} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

A candidate is admitted to further analysis if (4):

$$\prod_{j=1}^k r_j(\mathbf{x}_i) = 1 \quad (4)$$

Otherwise, the candidate is classified into class $y_i = 0$ ("rejected candidate").

The set of data proceeding to the next layer can be denoted as (5):

$$\mathcal{X}^{(1)} = \{\mathbf{x}_i: \prod_{j=1}^k r_j(\mathbf{x}_i) = 1\}. \quad (5)$$

Candidates who meet the requirements proceed to the next layer. This stage is responsible for reducing the problem space and ensuring the model's compliance with expert domain knowledge.

Step 3. Implementation of the Machine Learning Process (Second and Third Layer)

In this step, machine learning algorithms are applied to the subset of candidates who meet the minimum rule-based requirements. The purpose of this application is to assign candidates to one of two clusters: the green cluster (invited to interview) and the yellow cluster (reserve candidates). To ensure a reliable assessment of model performance and to limit the impact of random data partitioning on results, it is proposed to use the Repeated K-Fold Cross-Validation procedure. This method involves repeatedly and randomly applying the standard k-fold cross-validation [23]. For instance, a configuration of $k = 10$ (ten folds) and $n = 10$ repetitions may be used, which results in a total of 100 independent validation iterations. The cross-validation procedure consists of the following stages:

Stage 1. Random Data Partitioning (First Level – “Folds”)

For each validation repetition:

1. The full dataset is randomly permuted.
2. The data is then divided into, for example, ten mutually disjoint and equally sized subsets.
3. In each of the ten iterations, one subset serves as the validation set, while the remaining nine form the training set.

In this way, ten folds are obtained within a single repetition.

Stage 2. Repetition of the Procedure (Second Level – “Repeats”)

The process described in Stage 1 is repeated, e.g., ten times, each time initializing a new random permutation of the dataset. These repetitions are intended to ensure that:

- each element of the dataset appears in the validation set multiple times,
- variability resulting from random data partitioning is significantly reduced,
- model performance estimates are more stable and reliable.

As a result of carrying out Stages 2 and 3, a total of 100 distinct training-validation configurations are obtained.

Stage 3. Data Scaling Within Folds

To avoid information leakage, feature normalization must be performed exclusively on the training set of each fold:

1. The training set can be scaled using methods such as Standard Scaling or Min-Max Scaling.
2. The scaler parameters are then applied to transform the validation set.

This ensures that the validation process adheres fully to the principles of a fair experiment.

Stage 4. Clustering Within the Fold and Selection of the Best Algorithm

Within each fold, on the normalized training set, the following is performed:

1. Independent fitting of unsupervised clustering algorithms, such as:
 - K-Means,
 - Gaussian Mixture Model,
 - Agglomerative Clustering.

The K-Means algorithm is one of the most commonly used unsupervised clustering methods, aimed at partitioning a set of observations into K disjoint clusters by minimizing within-cluster variance [24]. Each cluster is represented by a centroid, which is the arithmetic mean of the points assigned to it. The algorithm operates iteratively, alternately assigning observations to the nearest centroid and updating the centroids' positions until convergence is achieved. The objective function of K-Means takes the form (6):

$$\min_{C_1, \dots, C_K} \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \quad (6)$$

where C_k denotes the set of observations in cluster k , and μ_k is the centroid vector of that cluster.

The Gaussian Mixture Model, on the other hand, is a probabilistic approach to clustering, in which it is assumed that the data originate from a mixture of K multivariate normal distributions. Unlike K-Means, GMM allows for soft assignment of observations to clusters by estimating

membership probabilities [25]. The model parameters (means, covariance matrices, and component weights) are typically estimated using the EM (Expectation–Maximization) algorithm. The mixture density function takes the form (7):

$$p(x_i) = \sum_{k=1}^K \pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k) \quad (7)$$

where π_k denotes the weight of the k -th component ($\sum_k \pi_k = 1$), and $\mathcal{N}(\cdot)$ is the normal distribution with mean μ_k and covariance matrix Σ_k .

Agglomerative Clustering, in turn, is a bottom-up hierarchical clustering method in which each observation is initially treated as a separate cluster. Then, in subsequent iterations, the most similar clusters are merged until a predefined number of clusters is reached or a complete hierarchy in the form of a dendrogram is obtained. A key component of the algorithm is the definition of the distance measure between clusters (the so-called linkage), such as single, complete, or average linkage [26]. For example, in the case of average linkage, the distance between two clusters A and B is defined as (8):

$$d(A, B) = \frac{1}{|A||B|} \sum_{x_i \in A} \sum_{x_j \in B} \|x_i - x_j\| \quad (8)$$

1. For each algorithm, cluster labels are generated only for the training data.
2. Based on these labels, the silhouette coefficient is calculated (also only on the training data).
3. The algorithm with the highest silhouette value is selected as the optimal clustering method for the given fold.

As a result, different folds could employ different unsupervised algorithms, depending on their local fit quality.

Stage 5. Generating pseudo-labels for the validation set

The selected clustering algorithm is then used to assign pseudo-labels to the validation set (classification into two clusters):

- if the algorithm had a prediction mechanism (e.g., K-Means, GMM), its `.predict()` method was used;
- pseudo-labels using the Agglomerative algorithm were determined based on the nearest centroid of the clusters formed in the training set.

In this way, consistent mapping of clusters to validation data was ensured without using any information from outside the fold.

Stage 6. Training and Evaluation of Supervised Classifiers

For each iteration:

1. Supervised classification models are trained on:
 - normalized training data,
 - pseudo-labels obtained from clustering.
2. The models are then evaluated on the validation set, using the pseudo-labels assigned to the test data of the fold.

In the proposed model, it is suggested to employ supervised learning models such as the Naive Bayes classifier, Support Vector Machines (both linear and nonlinear), Decision Trees, the k-NN algorithm, Logistic Regression, Random Forest, and Gradient Boosting.

The Naive Bayes classifier is a probabilistic classification model based on Bayes' theorem and the assumption of conditional independence of features with respect to the decision class. Despite this strong simplifying assumption, the algorithm often achieves high performance, especially in tasks involving a large number of features and a limited number of observations [27]. In the recruitment context, it enables fast and interpretable estimation of the probability that a candidate belongs to a given decision class. The decision rule takes the form (9):

$$\hat{y} = \operatorname{argmax}_{y \in \mathcal{Y}} P(y) \prod_{j=1}^d P(x_j|y) \quad (9)$$

Support Vector Machines, on the other hand, are classification methods based on the concept of maximizing the margin between classes [28]. In the linear case, SVM determines a decision hyperplane in the feature space, whereas the nonlinear variant employs a kernel function (the kernel trick), allowing for data separation in a higher-dimensional space. SVM is characterized by high robustness to overfitting and strong performance in high-dimensional spaces. The optimization problem for SVM takes the form (10):

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (10)$$

subject to the constraints (11):

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad (11)$$

Decision Trees are classification models based on recursively partitioning the feature space according to if-then rules. Each node of the tree represents a decision based on the value of a single feature, and the leaves correspond to decision classes [29]. These models are highly interpretable, which makes them particularly useful in regulated contexts such as recruitment. A typical splitting criterion is the minimization of impurity (e.g., the Gini index) (12):

$$G(t) = 1 - \sum_{k=1}^K p_k^2 \quad (12)$$

where p_k denotes the proportion of class k in node t .

The kNN algorithm, in turn, is a non-parametric classification method based on the assumption that observations similar to each other in the feature space belong to the same class [30]. The classification decision for a new observation is made based on the classes of the k nearest neighbors in the training set, using a selected distance metric. The decision rule is given by (13):

$$\hat{y} = \operatorname{mode}\{y_i: x_i \in \mathcal{N}_k(x)\} \quad (13)$$

where $\mathcal{N}_k(x)$ denotes the set of k nearest neighbors of observation x .

Logistic regression is a linear probabilistic model used for binary classification, which models the probability of belonging to the positive class using the logistic function [31]. Despite its simple structure, this model offers high interpretability of coefficients and stability of estimation, which makes it a commonly used benchmark in comparative analyses. The prediction function is given by (14):

$$P(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta^T x)}} \quad (14)$$

Random Forest is an ensemble machine learning algorithm that combines the predictions of multiple decision trees trained on random subsamples of the data and random subsets of features. This mechanism reduces model variance and improves its generalization ability [32]. Additionally, Random Forest allows for the estimation of feature importance, which is crucial from the interpretability perspective. The prediction of the tree ensemble is given by (15):

$$\hat{y} = \operatorname{mode}\{h_1(x), h_2(x), \dots, h_M(x)\} \quad (15)$$

where $h_m(x)$ denotes the prediction of the m -th tree.

Gradient Boosting is an ensemble method based on the sequential training of weak models (most often decision trees), where each subsequent model corrects the errors made by the previous ones [33]. The training process is carried out by minimizing the loss function using the gradient method

in the function space. This algorithm is characterized by high predictive performance, especially in complex classification problems. The general form of the model is given by (16):

$$F_M(x) = \sum_{m=1}^M \gamma_m h_m(x) \quad (16)$$

where $h_m(x)$ denotes a weak classifier, and γ_m is its weight determined during the optimization process.

Stage 7. Result Aggregation

After completing 100 folds:

- for each classification model, the mean and standard deviation of accuracy are calculated,
- the frequency of each clustering algorithm being selected as the “best” in a given fold is analyzed,
- final clustering is performed on the full dataset using the globally selected best algorithm.

The models should be evaluated using standard performance metrics such as accuracy, precision, recall, and F1-score. The model selected should demonstrate the highest predictive stability and generalization capability. This layer constitutes the operational component of the system-it performs real-time classification for new candidates.

Step 4. Analysis of the Influence of Individual Features on Cluster Assignment (XAI Layer)

This step investigates which variables have a significant impact on assigning candidates to clusters. The goal of the XAI layer is to ensure transparency and interpretability of the decision-making process during recruitment. This is particularly important as candidate selection also carries ethical and legal implications. To assess the impact of individual features on cluster assignment, the SHAP method is proposed. It is based on Shapley values from cooperative game theory. The SHAP value describes the contribution of a given feature to the prediction by averaging its marginal contribution across all possible combinations of the remaining features. The Shapley value for feature i in this method is computed using formula (17).

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|! \cdot (|F| - |S| - 1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)] \quad (17)$$

where:

- ϕ_i - SHAP value for feature i
- F - the set of all features
- S - a subset of features not containing feature i
- $|S|$ - the number of elements in set S
- $|F|$ - the total number of features
- $f_S(x_S)$ - the model prediction using only the features in set S
- $f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)$ - the model prediction using the features in set S plus feature i
- x_S - values of the features in set S

SHAP is considered one of the most reliable and theoretically grounded methods for explaining machine learning models [34]. Through the integration of interpretable XAI methods, the model remains transparent, compliant with AI ethics principles, and enables decision auditability.

Step 5. Validation of the Obtained Clusters Using the Necessity Condition Analysis (NCA) Method

As a result of the subsequent steps of the model, candidates have been classified, particularly into the green and yellow groups. Such clustering allows us to conclude that candidates in the green cluster are better than those in the yellow cluster. However, it cannot be ruled out that both clusters are weak (below the recruiter's expectations) or that both are strong (above expectations). To provide the recruiter with comprehensive information on cluster quality, an additional validation step is required. The developed model proposes using the NCA method for this validation. Necessary Condition Analysis (NCA) is a method for identifying so-called necessary conditions for achieving a

certain level of outcome [35]. Unlike traditional correlational and regression methods, which assess average effects (“the more X, the more Y on average”), NCA focuses on the logical necessity relationship—that is, it checks whether the absence of a certain level of X prevents achieving a high level of Y. Formally, X is a necessary condition for Y if high values of Y occur only when X exceeds a certain minimum threshold x^* . One of the practical algorithms used in NCA is CE–FDH (Ceiling Envelopment – Free Disposal Hull) [35]. This method constructs a so-called ceiling line over the cloud of points (x_i, y_i) , representing observations. For the increasing ordered values of x , it computes the maximum value of y achieved so far (18):

$$\text{ceiling}(x_i) = \max_{j \leq i} (y_j) \quad (18)$$

In this way, a monotonic curve is formed, representing the upper boundary of the observed outcomes. Then, for a given target level y_{target} , the so-called bottleneck point x^* is calculated—this is the smallest value of variable X for which the “ceiling” exceeds the level y_{target} (19).

$$x^* = \min \{x_i : \text{ceiling}(x_i) \geq y_{\text{target}}\} \quad (19)$$

Interpretatively, x^* represents the minimum level of the necessary condition that must be met in order to achieve the desired outcome y_{target} . When applied to candidate evaluation, this method enables the identification of the minimum competence threshold (X) required to attain a satisfactory level of overall quality (Y). As a result, beyond merely classifying candidates into clusters (“Green”/“Yellow”), it becomes possible to further differentiate them into four categories, e.g., Green Above, Green Below, Yellow Above, Yellow Below—depending on whether they meet the necessary condition determined by the NCA analysis.

4. Results

To evaluate the proposed model, a case study was conducted in which potential candidates for the position of Data Scientist completed a prepared application form. This form was developed by a Human Resource Management expert and reflects the key areas of knowledge and skills required for the specified position. The questionnaire was published on November 17, 2025, and data collection was concluded on November 24, 2025. It should be emphasized that the dataset used serves an illustrative purpose—its goal is to demonstrate the practical operation of the model and the interaction between subsequent layers of the algorithm, rather than to assess the model’s effectiveness in terms of general, multi-contextual validation. Thus, the collected data enables the demonstration of the model’s architecture and the rationale behind decision-making, in accordance with the concept presented in the Methods section.

Step 1. Development of Evaluation Criteria for Candidate Application Forms

Table 1 presents the evaluation criteria proposed by the recruiter for the Data Scientist position, along with the corresponding rating scales.

Table 1. Structure of the Application Form for the Data Scientist Position.

No	Criterion	Criterion Evaluation Scale				
f1	Education	primary	secondary	post-secondary	higher (bachelor/engineer)	higher (master's) or PhD
f2	Completion of a degree program related to Data Science (Data Science, Computer Science, Artificial Intelligence, Mathematics)				yes	no
f3	Experience in Data Science or related fields					Number of years
f4	Total professional experience					Number of years
f5	Knowledge of Python	1	2	3	4	5
f6	Knowledge of R	1	2	3	4	5
f7	Knowledge of SQL	1	2	3	4	5
f8	Knowledge of NoSQL databases (MongoDB, Cassandra)					

f9	Knowledge of ML tools (scikit-learn, TensorFlow, Pytorch)	1	2	3	4	5				
f10	Knowledge of statistics	1	2	3	4	5				
f11	Dashboard creation skills (PowerBi, Tableau)	1	2	3	4	5				
f12	Knowledge of Big Data technologies (Spark, Hadoop)	1	2	3	4	5				
f13	Knowledge of cloud tools (AWS, Azure)	1	2	3	4	5				
f14	Knowledge of machine learning algorithms	1	2	3	4	5				
f15	English language proficiency	1	2	3	4	5				
f16	Knowledge of version control systems (e.g. GIT)	1	2	3	4	5				
f17	Knowledge of MLOps/CI-CD (Mlflow, Docker, Airflow)	1	2	3	4	5				
f18	Published scientific articles						yes		no	

As a result of the recruitment process, 163 candidates applied and completed the described form.

Step 2. Determining the minimum values for each evaluation criterion by the recruiter to allow an invitation to a job interview (rule layer – preliminary filtering)

In the second step, the recruiter formulated a list of minimum required values for each field in the completed form. This list is presented in Table 2.

Table 2. List of minimum required values for individual evaluation criteria in the recruitment for the position of Data Scientist.

Criterion	f1	f2	f3	f4	f5	f6	f7	f8	f9
Minimum value	Higher (bachelor/engineer)	YES	2 years	2 years	3	1	2	1	2
Criterion	f10	f11	f12	f13	f14	f15	f16	f17	f18
Minimum value	3	1	1	1	2	2	2	1	YES

It turned out that 77 candidates did not meet the minimum requirements specified by the recruiter for the Data Scientist position. This means that these candidates are classified into the red cluster, which will not be invited to a job interview. The remaining candidates proceed to the next stages of the developed model procedure.

Step 3. Implementation of the machine learning process (second and third layer)

As part of this step, an unsupervised machine learning process was carried out using cross-validation (10 folds and 10 repetitions) for the following algorithms: K-Means, Gaussian Mixture Model, and Agglomerative Clustering. The purpose of this step was to assign pseudo-labels to the candidates in the training sets into two clusters (green and yellow). For each fold, labels were generated using the three mentioned algorithms, and based on the silhouette coefficient (calculated only for the training data), the cluster indicated by the most effective algorithm was selected. Naturally, in different iterations, different unsupervised learning algorithms turned out to be the most effective. The distribution of algorithm selection results is presented in Table 3.

Table 3. Distribution of unsupervised learning algorithm selection results.

Algorithm	Number of iterations in which the algorithm was the most effective based on the silhouette coefficient
K-Means	96
Gaussian Mixture Model	4
Agglomerative Clustering	0

The presented results clearly indicate that the K-Means algorithm was the most effective in the vast majority of validation iterations-achieving the highest silhouette score in 96 out of 100 configurations. The Gaussian Mixture Model proved to be the best in only 4 iterations, while Agglomerative Clustering was not selected even once as the optimal method. Such a clear dominance of K-Means suggests that the structure of the candidate data is best represented by a centroid-based approach, and the feature space forms relatively distinct, spherical clusters. In each iteration, after determining the pseudo-labels, supervised machine learning algorithms were implemented. The

employed machine learning algorithms included: Naive Bayes classifier, Support Vector Machines (both linear and nonlinear), decision trees, the k-NN algorithm, logistic regression, random forest, and Gradient Boosting. These models were then evaluated on the validation set using the pseudo-labels assigned to the test data of the fold. Table 4 presents the results of supervised learning for the analyzed algorithms.

Table 4. Results of the supervised machine learning process in the cross-validation model.

Algorithm	Average Accuracy	Standard Deviation of Accuracy
Naive Bayes Classifier	0.7933	0.23
Linear Support Vector Machine	0.9750	0.09
Nonlinear Support Vector Machine	0.9717	0.09
Decision Trees	0.8475	0.19
k-NN Algorithm	0.9525	0.11
Logistic Regression	0.9883	0.06
Random Forests	0.9517	0.11
Gradient Boosting	0.8633	0.20

The comparison of results obtained by the analyzed classifiers indicates that logistic regression demonstrated the highest predictive accuracy, achieving an average accuracy of 0.9883 with the lowest standard deviation (0.06). Such high stability suggests that this model generalizes particularly well to the pseudo-label structure derived from clustering. Very good results were also achieved by the linear SVM (0.9750) and nonlinear SVM (0.9717), confirming the high separability of classes in the feature space and the relative linearity of the decision boundary. Models with more complex structures, such as Random Forests (0.9517) and k-NN (0.9525), achieved moderately high results, but their stability was lower compared to logistic regression and SVM. The weakest performance was recorded by Decision Trees (0.8475) and Naive Bayes (0.7933), indicating that these models may not be fully suitable for the characteristics of the candidate data. Figure 1 presents the accuracy distribution of all models.

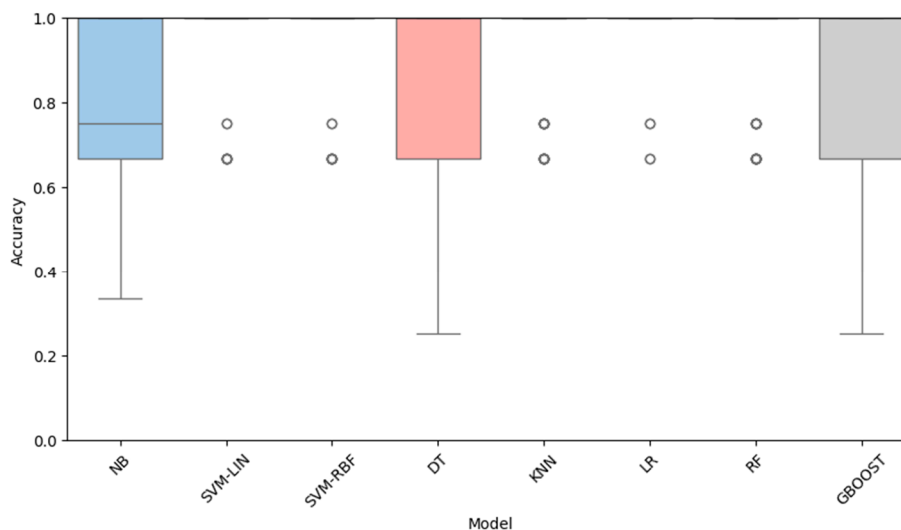


Figure 1. Accuracy distribution of supervised machine learning models.

The chart clearly shows the dominance of logistic regression and linear SVM, which achieve the highest median accuracy with relatively low variability in results. The k-NN, Random Forest, and SVM-RBF models also exhibit high accuracy, though with slightly greater dispersion. In contrast, Naive Bayes records significantly lower accuracy values, while decision trees and Gradient Boosting display higher variability, indicating less stable performance compared to the linear models. Figure 2 presents the histogram of the best-performing algorithm (logistic regression).

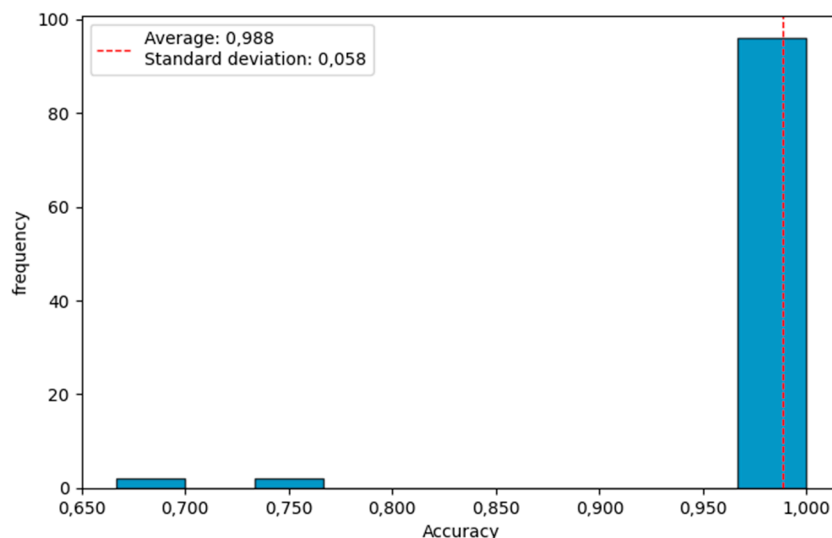


Figure 2. Histogram of the most effective supervised learning algorithm.

The histogram shows that logistic regression achieves very high and exceptionally stable accuracy. The vast majority of accuracy values are concentrated within the narrow range of 0.95–1.00, confirming both the high effectiveness and low variability of the model. Only a few observations fall outside this range, but their number is marginal. Such a concentrated distribution indicates that logistic regression is the most reliable and consistent model in the analyzed validation procedure.

Step 4. Analysis of the impact of individual features on cluster assignment (XAI layer)

In this step, the influence of individual candidate features on the assignment to one of the two clusters-green and yellow-is examined. To carry out this task, the SHAP method was used. Its results are presented in Figure 3.

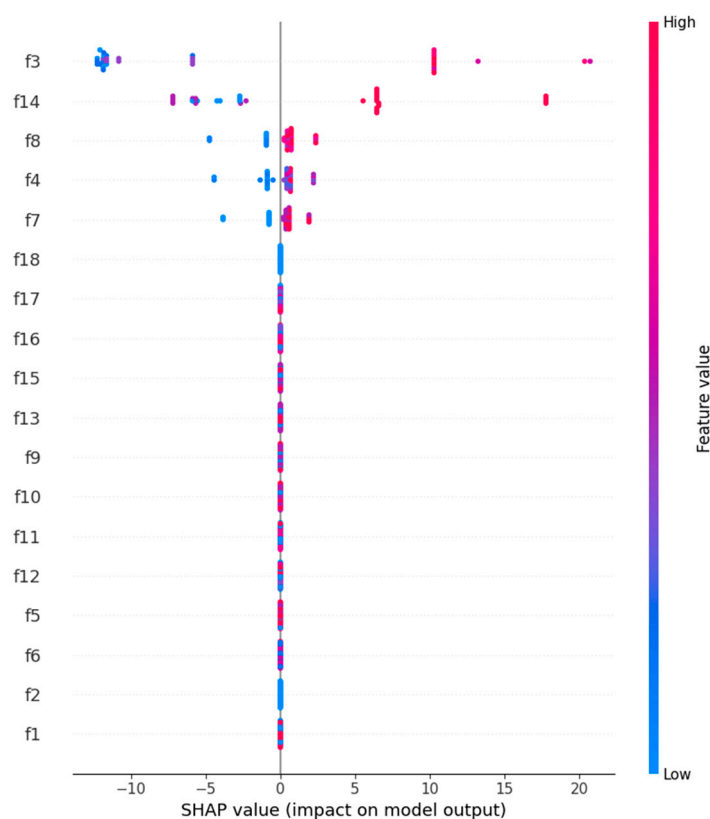


Figure 3. Results of the SHAP analysis.

The SHAP analysis results indicate that among candidates who passed the minimum requirements layer, the most influential features for assignment to the green cluster are competencies most directly related to performing Data Scientist tasks-such as knowledge of machine learning algorithms (f14), experience in Data Science (f3), overall professional experience (f4), and skills related to NoSQL database technologies (f8). High values of these features clearly increase the likelihood of being classified into the group of recommended candidates, as shown by red points on the positive side of the SHAP axis. At the same time, variables such as knowledge of Python (f5), SQL (f7), or statistics (f10) have a smaller impact on distinguishing between clusters-not because they are unimportant, but because the initial filtering eliminated all individuals with low levels of these competencies, leaving a relatively homogeneous group in which differences in these skills are already much smaller. As a result, SHAP emphasizes more strongly those features that differentiate the top candidates from each other, rather than those that were necessary conditions for passing the first stage of selection.

Step 5. Validation of the obtained clusters using the Necessity Condition Analysis (NCA) method

In the final stage of the model, the quality of the obtained clusters was validated using NCA analysis, which aims to assess whether the competency level of candidates in a given cluster meets the minimum threshold required to achieve a satisfactory level of overall quality. As outlined in the Methods section, NCA allows for determining whether a specific outcome level (Y) can be attained only if the candidate meets a minimum level of the most critical competencies (X). In the empirical analysis, the measure of overall quality was defined as the Y_score, calculated as the average value of all candidate features (f1–f18), while the level of the necessary condition, X_cond, was defined as the minimum of these features, representing the weakest element of the competency profile. For the assumed target level $Y_{target} = 0.60$, the CE–FDH algorithm determined the bottleneck value $X = -0.483^*$. This means that for a candidate to achieve the desired level of overall quality, their weakest competency must exceed a defined minimum threshold. Next, each candidate was evaluated against two conditions: (1) achieving a quality level not lower than Y_{target} , and (2) having X_{cond} not less than X^* . Candidates meeting both criteria were labeled as Above, and the remaining ones as Below. The NCA results were then combined with the earlier cluster classification (Green/Yellow), resulting in the aggregation of candidates into four quality categories: Green (Above), Green (Below), Yellow (Above), and Yellow (Below). The following distribution of groups was obtained in the analyzed dataset:

- Green (Above): 10 candidates
- Green (Below): 29 candidates
- Yellow (Above): 0 candidates
- Yellow (Below): 47 candidates

The results clearly confirm that although the clustering model effectively separates a group of clearly stronger (Green) and weaker (Yellow) candidates, not all individuals assigned to the Green cluster meet the necessary condition defined by NCA. Only a small portion of those in the Green cluster (10 individuals) exceed both the quality threshold (Y_{target}) and the minimum competency threshold (X^*). This indicates that the Green segment is internally diverse-besides very strong candidates (Green Above), it also includes individuals with good overall performance who nevertheless have at least one distinctly weak competency (Green Below). In contrast, all candidates in the Yellow cluster were classified as Yellow (Below), which confirms the model's consistency-individuals previously identified as weaker do not meet the necessary condition and thus fail to reach even the minimal expected level of quality.

The application of NCA thus not only confirmed the distinction between the Green and Yellow clusters but also added an additional diagnostic layer, enabling precise evaluation of the "strong" candidates and identification of those who-despite a positive classification-do not meet the minimum threshold for the most essential competencies. In recruitment practice, this represents valuable

information for the recruiter, allowing for more informed differentiation of candidates by considering both their overall level and potential competency gaps.

5. Conclusions

The article presents a multi-layered, interpretable recruitment model integrating expert rules, unsupervised clustering, supervised classification models, and Explainable AI techniques, supplemented with validation using Necessary Condition Analysis. From a theoretical contribution perspective, the model offers an integrated approach to designing AI systems in HR, combining deterministic selection rules with probabilistic machine learning methods and formal tools for analyzing necessary conditions. The novelty of the solution lies in treating clustering as an intermediate layer that generates pseudo-labels for supervised models, enabling the combination of the flexibility of unsupervised methods with the high effectiveness of logistic regression and SVM, while maintaining interpretability at both the local feature level (SHAP) and in terms of the necessary conditions for a specified quality level (NCA). In the AI-in-HR literature, such a consistently formalized architecture is rare-where each layer serves a distinct function: minimum requirements filtering, structuring the candidate space, operational classification, and quality validation.

From a practical standpoint, the model can serve as real support for HR departments seeking solutions that are both efficient and compliant with growing ethical and regulatory demands, including the EU AI Act. The minimum rules layer reflects recruiter expertise and ensures alignment with employment policies. The use of cross-validation and a multi-model approach to clustering and classification reduces the risk of basing decisions on a single, randomly fitted model, enhancing recommendation stability. Empirical results show that logistic regression, in particular, exhibits very high accuracy with low variability, making it an attractive candidate for deployment. The XAI layer enables the generation of explanations understandable to practitioners at both the individual and global levels, while the application of NCA additionally differentiates candidates within clusters (e.g., Green Above vs. Green Below), enhancing the ability to manage recruitment risk in contexts of high application volume and limited time resources.

It is important to note that the presented example is illustrative in nature and intended solely to demonstrate the operation of the developed model, rather than to provide full-scale empirical validation. The dataset used, concerning recruitment for a Data Scientist position, serves as a case study enabling the demonstration of successive algorithm layers and their interactions. Thus, the aim was not to assess the model's universality across various industry contexts, but to show how the proposed architecture can be practically implemented. At the same time, the nature of the data-a relatively well-quantified competency profile-facilitates a clear presentation of the method's functioning. It must also be emphasized that the pseudo-labels from the clustering layer serve as a structural element of the model rather than a reference benchmark of quality, and the interpretations generated by SHAP and NCA illustrate the interpretability capabilities of the architecture rather than an assessment of the fairness of the recruitment process. Consequently, the presented example should be treated as a proof of concept, while full-scale, multi-context empirical validation remains a separate area for future research.

Therefore, future studies should focus on verifying the model in diverse organizational contexts and for a variety of job profiles, including less technical roles where data is more qualitative in nature. Particularly important is linking model results with data on employee performance, retention, and career development paths, which would allow for assessing predictive validity over a longer time horizon. Another promising direction is the development of user interfaces based on XAI, which would enable recruiters to explore SHAP and NCA outputs, simulate "what-if" scenarios, and easily document decisions in light of regulatory requirements. Finally, it is worth integrating explicitly defined fairness and non-discrimination metrics into the architecture so that fairness assessment becomes an integral part of validation alongside traditional performance measures. In this way, the proposed model could form the foundation for the next generation of ethical, transparent, and regulation-compliant AI systems in human resource management.

Author Contributions: Conceptualization, M.N. and M.P.N.; methodology, M.N.; software, M.N.; validation, M.N. and M.P.N.; formal analysis, M.N. and M.P.N.; investigation, M.N. and M.P.N.; resources, M.N. and M.P.N.; data curation, M.N.; writing—original draft preparation, M.N. and M.P.N.; writing—review and editing, M.N. and M.P.N.; visualization, M.N.; supervision, M.N. and M.P.N.; project administration, M.N.; funding acquisition, M.N. and M.P.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science Centre, Poland, under grant no. 2024/08/X/HS4/00155 (project leader: Marcin Nowak), and by project no. 0813/SBAD/2991 entitled “The Use of Artificial Intelligence in Management 4.0: Modeling Trust in Modern Technologies” (project leader: Marta Pawłowska-Nowak).

Institutional Review Board Statement: Ethical review and approval were waived for this study due to the fact that the research was conducted on a fully anonymous and voluntary basis. The data collection process did not involve sensitive personal information, medical data, or personally identifiable information (PII). The participants filled out the forms anonymously for the sole purpose of validating the proposed recruitment model, and the study did not pose any risk to their physical or psychological well-being.

Data Availability Statement: Marcin Nowak has unlimited and free-of-charge access to the dataset.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Faroozan, A. The Evolving Role of Artificial Intelligence in Recruitment: Efficiency, Bias Mitigation, and Ethical Challenges. *International Journal For Multidisciplinary Research* **2025**, *7*.
2. HRCI History of AI in HR 2025. Available online: <https://www.hrci.org/community/blogs-and-announcements/hr-leads-business-blog/hr-leads-business/2025/08/14/history-of-ai-in-hr#:~:text=In%20the%201980s%20and%201990s%2C,could%20achieve%20in%20human%20resources> (accessed on 01.12.2025).
3. Fabris, A.; Baranowska, N.; Dennis, M.J.; Graus, D.; Hacker, P.; Saldivar, J.; Zuiderveen Borgesius, F.; Biega, A.J. Fairness and Bias in Algorithmic Hiring: A Multidisciplinary Survey. *ACM Trans Intell Syst Technol* **2025**, *16*, 1–54.
4. Lavanchy, M.; Reichert, P.; Narayanan, J.; Savani, K. Applicants’ Fairness Perceptions of Algorithm-Driven Hiring Procedures. *Journal of Business Ethics* **2023**, *188*, 125–150.
5. Liu, Q.; Wan, H.; Yu, H. The Application of Deep Learning in Human Resource Management: A New Perspective on Employee Recruitment and Performance Evaluation. *Academic Journal of Management and Social Sciences* **2023**, *3*, 101–104.
6. Hemalatha, A.; Kumari, P.B.; Nawaz, N.; Gajenderan, V. Impact of Artificial Intelligence on Recruitment and Selection of Information Technology Companies. In Proceedings of the 2021 international conference on artificial intelligence and smart systems (ICAIS); **2021**; pp. 60–66.
7. Madanchian, M. From Recruitment to Retention: AI Tools for Human Resource Decision-Making. *Applied Sciences* **2024**, *14*, 11750.
8. Ali, A.; Rafi, N. Enhancing Human Resource Management Through Advanced Decision-Making Strategies: Harnessing The Power Of Artificial Intelligence For Strategic, Data-Driven, And Judicious Choices. *Journal of Human Resource Management* **2024**, *21*, 881–889.
9. HireBee AI in HR Statistics **2024**. Available online: <https://hirebee.ai/blog/ai-in-hr-statistics/#:~:text=1.%2045,AI%20adoption%20rate%20for%20HRM> (accessed on 01.12.2025).
10. Zhang, G.; Pan, L.; Tang, F.; Yao, F. Explainable Artificial Intelligence in the Talent Recruitment Process—a Literature Review. *Cogent Business & Management* **2025**, *12*, 2570881.
11. Chen, Z. Ethics and Discrimination in Artificial Intelligence-Enabled Recruitment Practices. *Humanit Soc Sci Commun* **2023**, *10*, 1–12.

12. Agbasiere, C.L.; Nze-Igwe, G.R. Algorithmic Fairness in Recruitment: Designing AI-Powered Hiring Tools to Identify and Reduce Biases in Candidate Selection. *Path of Science* **2025**, *11*, 5001–5021.
13. Dastin, J. Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women. *Reuters* **2018**.
14. Hunkenschroer, A.L.; Kriebitz, A. Is AI Recruiting (Un) Ethical? A Human Rights Perspective on the Use of AI for Hiring. *AI and Ethics* **2023**, *3*, 199–213.
15. Society for Human Resource Management Fresh SHRM Research Explores Use of Automation and AI in HR 2024.
16. Sykorová, Z.; Hague, D.; Dvoulet, O.; Procházka, D.A. Incorporating Artificial Intelligence (AI) into Recruitment Processes: Ethical Considerations. *Vilakshan-XIMB Journal of Management* **2024**.
17. Qiang, R.E.N.; Jing, D.U. Harmonizing Innovation and Regulation: The EU Artificial Intelligence Act in the International Trade Context. *Computer Law & Security Review* **2024**, *54*, 106028.
18. Fisher Phillips European Industry Pushes Back on the EU AI Act Practical Impact 2024. . Available online: <https://www.fisherphillips.com/en/news-insights/european-industry-pushes-back-on-the-eu-ai-act.html#:~:text=Practical%20Impact> (accessed on 01.12.2025).
19. Marin Diaz, G.; Galán Hernández, J.J.; Galdón Salvador, J.L. Analyzing Employee Attrition Using Explainable AI for Strategic HR Decision-Making. *Mathematics* **2023**, *11*, 4677.
20. Thalpage, N. Unlocking the Black Box: Explainable Artificial Intelligence (XAI) for Trust and Transparency in Ai Systems. *J. Digit. Art Humanit* **2023**, *4*, 31–36.
21. Pinto, G.B.S.; Mello, C.E.; Garcia, A.C.B. Explainable AI in Labor Market Applications **2025**. Available online: <https://www.scitepress.org/Papers/2025/133841/133841.pdf> (accessed on 01.12.2025).
22. Nowak, M.; Rabczun, A.; Łopatka, P. Impact of Electrification on African Development-Analysis with Using Grey Systems Theory. *Energies (Basel)* **2021**, *14*, 5181.
23. Kaliappan, J.; Bagepalli, A.R.; Almal, S.; Mishra, R.; Hu, Y.-C.; Srinivasan, K. Impact of Cross-Validation on Machine Learning Models for Early Detection of Intrauterine Fetal Demise. *Diagnostics* **2023**, *13*, 1692.
24. Ufeli, C.P.; Sattar, M.U.; Hasan, R.; Mahmood, S. Enhancing Customer Segmentation Through Factor Analysis of Mixed Data (FAMD)-Based Approach Using K-Means and Hierarchical Clustering Algorithms. *Information* **2025**, *16*, 441.
25. Lodygowski, T.; Szrama, S. Unsupervised Classification and Remaining Useful Life Prediction for Turbofan Engines Using Autoencoders and Gaussian Mixture Models: A Comprehensive Framework for Predictive Maintenance. *Applied Sciences* **2025**, *15*, 7884.
26. Contreras, J.M.; Molina Portillo, E.; Fernández Luna, J.M. Evaluation of Hierarchical Clustering Methodologies for Identifying Patterns in Timeout Requests in EuroLeague Basketball. *Mathematics* **2025**, *13*, 2414.
27. Phatcharathada, B.; Srisuradetchai, P. Randomized Feature and Bootstrapped Naive Bayes Classification. *Applied System Innovation* **2025**, *8*, 94.
28. Nowak, M.; Pawłowska-Nowak, M. Dynamic Pricing Method in the E-Commerce Industry Using Machine Learning. *Applied Sciences (2076-3417)* **2024**, *14*.
29. Gajowniczek, K.; Zabkowski, T. Interactive Decision Tree Learning and Decision Rule Extraction Based on the ImbTreeEntropy and ImbTreeAUC Packages. *Processes* **2021**, *9*, 1107.
30. Nowak, M.; Zajkowski, R. An Integrated Structural Equation Modelling and Machine Learning Framework for Measurement Scale Evaluation—Application to Voluntary Turnover Intentions. *AppliedMath* **2025**, *5*, 105.
31. Nikolić, M.; Nikolić, D.; Stefanović, M.; Koprivica, S.; Stefanović, D. Mitigating Algorithmic Bias Through Probability Calibration: A Case Study on Lead Generation Data. *Mathematics* **2025**, *13*, 2183.
32. Amamra, S.-A. Random Forest-Based Machine Learning Model Design for 21,700/5 Ah Lithium Cell Health Prediction Using Experimental Data. *Physchem* **2025**, *5*, 12.
33. Airlangga, G.; Liu, A. A Hybrid Gradient Boosting and Neural Network Model for Predicting Urban Happiness: Integrating Ensemble Learning with Deep Representation for Enhanced Accuracy. *Mach Learn Knowl Extr* **2025**, *7*, 4.

34. Mosca, E.; Szigeti, F.; Tragianni, S.; Gallagher, D.; Groh, G. SHAP-Based Explanation Methods: A Review for NLP Interpretability. In Proceedings of the Proceedings of the 29th international conference on computational linguistics; **2022**; pp. 4593–4603.
35. Dul, J.; Hauff, S.; Bouncken, R.B. Necessary Condition Analysis (NCA): Review of Research Topics and Guidelines for Good Practice. *Review of Managerial Science* **2023**, *17*, 683–714.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.