ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FUZZY ADAPTIVE DENSITY WEIGHTED K MEANS FRAMEWORK FOR RECRUITMENT CANDIDATE CLASSIFICATION USING DENSITY BASED CLUSTERING AND WEIGHTED SIMILARITY ANALYSIS

PETRICIA LEEMA ROSELINE[1], ARUNA P[2]

[1]Research Scholar, Periyar Maniammai Institute of Science & Technology, Department of Computer Applications, Thanjavur, Tamil Nadu, India
[2] Periyar Maniammai Institute of Science & Technology, Department of Computer Applications, Thanjavur, Tamil Nadu, India
E-mail: [1]leema.r.mohan@gmail.com, [2]aruna_se@pmu.edu

ABSTRACT

Modern recruitment environments generate large volumes of candidate data that require structured analytical mechanisms for effective competency evaluation and decision support. Conventional recruitment analytics approaches mainly emphasize recommendation systems or transparency-oriented analysis, offering limited capability for discovering competency structures within candidate datasets. Effective recruitment intelligence requires analytical frameworks capable of identifying meaningful candidate groupings while preserving variations in skill composition and experience distribution. This study introduces a Fuzzy-Adaptive Density-Weighted K-Means framework designed for competency-oriented clustering within recruitment data. The framework integrates adaptive density estimation, weighted similarity assessment, and fuzzy membership assignment to capture heterogeneous relationships among candidate attributes. Density-aware weighting strengthens recognition of candidate distribution patterns, while fuzzy membership modelling enables flexible representation of overlapping competency profiles. Iterative centroid refinement further supports stable cluster formation and improved structural consistency. Experimental evaluation conducted on a recruitment dataset containing ten thousand records demonstrates strong classification stability and reliable clustering behaviour. Analytical outcomes highlight the capability of the proposed framework to support structured candidate profiling and intelligent recruitment analytics within data-driven human resource environments.
Keywords: *Fuzzy Clustering, Recruitment Analytics, Density-Based Clustering, Candidate Competency Analysis, Weighted Similarity Measurement, Intelligent Recruitment Systems.*

1. INTRODUCTION

Intelligent recruitment analytics has emerged as a significant component within contemporary employment ecosystems that evaluate graduate applicants through structured data interpretation. Organizations collect extensive information related to candidate academic performance, technical proficiency, domain knowledge, communication capability, and interview evaluation. Recruitment environments generate large volumes of structured records obtained from application portals, testing platforms, and assessment systems [1]. Analytical techniques enable systematic interpretation of these datasets, revealing patterns that assist recruiters in understanding variations among candidate competencies. Data-oriented evaluation strengthens transparency and consistency within hiring processes by supporting objective interpretation of applicant attributes. Advanced computational models assist in discovering relationships among diverse candidate indicators, enabling deeper insight into graduate employability characteristics and improving analytical capabilities within recruitment management systems [2].

Candidate profiling represents an essential analytical activity in recruitment analytics where applicants are organized based on comparable competency attributes. Profiling approaches analyze educational qualifications, technical knowledge indicators, communication ratings, domain expertise, and evaluation scores obtained through recruitment assessments [3]. Structured representation of candidate attributes enables organized examination of applicant capabilities and facilitates meaningful interpretation of large recruitment datasets. Analytical profiling supports identification of groups that reflect varying levels of professional readiness among graduate applicants. Recruitment analysts rely on such structured grouping mechanisms to understand competency distributions across diverse academic and technical backgrounds. Organized candidate profiles enhance interpretability of applicant records and contribute toward systematic evaluation environments that support recruitment decision processes [4].

Clustering techniques form an important category of unsupervised learning methods used for grouping observations that share similar characteristics within multidimensional datasets. These analytical models operate without predefined labels and identify structural relationships through similarity measurement across attributes [5]. Recruitment datasets contain numerous candidate indicators including educational achievements, skill assessments, interview evaluations, and communication scores. Clustering analysis enables discovery of hidden structures embedded within such records and supports identification of groups representing related competency patterns [6]. Analytical grouping generated through clustering contributes toward clearer interpretation of candidate attributes and assists recruitment analysts in exploring variations present across applicant populations. Unsupervised learning approaches provide valuable mechanisms for investigating patterns that may remain difficult to detect through conventional manual evaluation methods [7].

Recruitment datasets frequently exhibit heterogeneous distributions across candidate attributes such as skill ratings, assessment scores, experience indicators, and academic performance levels. Certain competencies appear densely represented across applicant populations, whereas other attributes occur less frequently within recruitment records. Analytical interpretation of such datasets requires mechanisms capable of identifying structural variations present within multidimensional feature spaces [8]. Density-oriented analysis assists in distinguishing concentrated candidate groups from regions containing sparse representation. Recognition of density characteristics supports more meaningful organization of applicant profiles and contributes toward clearer identification of competency clusters present within recruitment datasets. Analytical strategies that consider variations in data distribution improve the representation of complex candidate characteristics within recruitment analytics frameworks [9].

Fuzzy logic introduces a flexible analytical representation capable of capturing uncertainty and gradual similarity relationships present within real-world datasets. Candidate competencies frequently demonstrate overlapping characteristics where individuals possess multiple skill dimensions with varying levels of proficiency [10]. Strict grouping mechanisms that assign single-category membership may fail to represent such complexity adequately. Fuzzy representation enables partial association with multiple clusters, providing a more realistic interpretation of candidate similarity relationships. Adaptive clustering mechanisms capable of incorporating such flexible representations strengthen analytical exploration of recruitment datasets by accommodating diversity in candidate attributes [11]. Analytical environments that integrate clustering and adaptive representation techniques enhance interpretability of graduate employability patterns and contribute toward improved understanding of applicant competency distributions within modern recruitment analytics systems.

1.1. Problem Statement

Recruitment analytics involves evaluation of graduate applicants using diverse attributes such as academic scores, skill assessments, communication ratings, and interview performance indicators. Conventional analytical models struggle to represent overlapping competency patterns and irregular distributions present within candidate datasets. Rigid grouping mechanisms limit accurate interpretation of applicant similarity structures. An analytical approach capable of organizing multidimensional candidate attributes through adaptive and flexible clustering remains essential for improving interpretation of employability patterns within recruitment data environments.

1.2. Motivation

Rapid growth of digital recruitment platforms has produced extensive datasets describing candidate academic and professional characteristics. Manual interpretation of such multidimensional information restricts discovery of meaningful competency relationships among applicants. Intelligent analytical models capable of identifying structural patterns within recruitment data provide deeper insight into graduate employability characteristics. Development of advanced clustering strategies encourages systematic exploration of applicant attributes and strengthens analytical understanding of skill distributions within modern recruitment evaluation environments.

1.3. Objective

The primary objective involves development of an analytical framework that organizes graduate applicant attributes through adaptive clustering techniques. The research aims to examine candidate records composed of academic indicators, skill evaluations, and communication ratings to identify meaningful competency groupings. Analytical clustering seeks to represent structural relationships present within multidimensional recruitment datasets and support systematic interpretation of employability characteristics through intelligent grouping of candidate profiles.

1.4. Scope

The research scope focuses on analytical examination of structured recruitment datasets containing graduate applicant attributes related to education, technical competence, communication ability, and evaluation scores. Clustering-based analysis facilitates exploration of similarity patterns across candidate records and identification of competency groupings within multidimensional feature spaces. The study contributes toward development of recruitment analytics mechanisms that enhance understanding of employability characteristics within graduate applicant populations.

The study aims to develop a clustering-based recruitment analytics framework for competency-oriented candidate classification. Objectives include designing density-aware clustering, implementing adaptive attribute weighting, enabling fuzzy membership-based grouping, and evaluating performance using standard classification metrics.

The study introduces a clustering framework integrating density-aware analysis, adaptive attribute weighting, and fuzzy membership modelling for recruitment analytics. The framework captures heterogeneous candidate distributions and overlapping competency structures through flexible membership representation and density-sensitive similarity evaluation. This integration improves competency-oriented grouping and enhances classification consistency within recruitment datasets.

2. LITERATURE REVIEW

The “Bayesian Stack Insight” [12] study applied a stacked ensemble learning framework optimized through Bayesian hyperparameter tuning. Multiple base classifiers were combined, and their outputs aggregated using meta-learning to enhance prediction accuracy. Feature importance was extracted using explainable AI techniques to clarify model reasoning. “Invisible Threat Mapping” [13] methodology conducted a systematic review of Shadow AI deployment and security incidents. Case studies and incident logs were collected to analyze attack vectors, governance gaps, and mitigation strategies. A threat taxonomy was created, and comparative analysis measured exposure risk across industries. The “Expert ELSI Appraisal” [14] study used semi-structured interviews to evaluate synthetic data applications in medical AI. Methodological steps included interview guide development, participant recruitment, thematic coding, and cross-validation of responses. ELSI checklists were applied to categorize ethical, legal, and social considerations. Data analysis employed qualitative content analysis to extract recurring patterns and expert insights. Triangulation ensured reliability across stakeholder perspectives.

“AI Public Governance” [15] applied a structured framework integrating AI tools within public administration workflows. It utilized process-mapping techniques combined with predictive analytics to evaluate potential efficiency gains. Data from multiple administrative departments were standardized, processed, and analyzed using algorithmic simulations to model decision outcomes under varying scenarios. Risk assessment matrices quantified impacts of AI adoption, while stakeholder engagement models provided insight into organizational readiness. “Medical Diagnostics Defense” [16] involved stress-testing deep learning architectures against adversarial perturbations in medical imaging datasets. Models were subjected to crafted attack scenarios to assess vulnerability levels. Defensive strategies included gradient masking, adversarial training, and model ensembling to improve resilience. Evaluation metrics quantified classification stability, prediction confidence, and error rates under tampered inputs.

“Air Traffic AI Clarity” [17] adopted a mixed-methods approach combining explainable AI (XAI) outputs with human factors assessment. Simulated air traffic scenarios were used to evaluate controller interactions with XAI-generated recommendations. Cognitive workload was measured using validated scales, while user acceptance was captured through structured questionnaires. The methodology integrated behavioral logging, eye-tracking data, and interaction metrics to quantify interpretability effects. Statistical analyses correlated expertise levels with reliance on AI suggestions. "Ethical AI Progression" [18] employed a conceptual modeling methodology to map ethical considerations across AI development cycles. Multi-tiered evaluation frameworks incorporated normative principles, compliance standards, and decision-making autonomy metrics. Simulation-based testing assessed the impact of varying levels of algorithmic independence on ethical outcomes. Cross-generational scenario analysis compared early-stage designs with advanced autonomous iterations, capturing potential moral trade-offs.

"Recruitment LLM Integration" [19] Methodology combined retrieval-augmented generation (RAG) with large language models to enhance candidate evaluation. Structured and unstructured recruitment data were preprocessed, embedded, and indexed to enable fast retrieval during candidate scoring. LLMs generated contextual insights, supplemented by historical hiring patterns and competency mapping. "Job Offer Classifier" [20] Methodology leveraged BERT embeddings in combination with O*NET occupational taxonomies for precise job classification. Textual data from job postings were tokenized, encoded, and mapped to standardized occupational categories. Feature extraction incorporated semantic similarity measures, while supervised learning models performed multi-class classification. Training and evaluation involved cross-validation and error analysis to optimize label assignment accuracy. "Bias Reduction Tuning" [21] applied adversarial contrastive learning to fine-tune large language models for resume screening. Candidate datasets were augmented with synthetically generated variations to expose potential bias patterns. Training optimized contrastive loss functions to align representations while minimizing group-specific disparities.

"Ethical Recruitment Blueprint" [22] developed a governance framework integrating ethical, fairness, and sustainability considerations into recruitment AI systems. Multi-dimensional compliance matrices evaluated algorithmic decisions, incorporating fairness audits, risk scoring, and sustainability impact assessment. Policy-driven simulations tested the framework's ability to guide ethical choices under varying recruitment scenarios. Structured stakeholder review cycles captured organizational and societal expectations. "Semantic Job Matching" [23] integrated semantic similarity algorithms with AI-driven frameworks to enhance job-candidate alignment. Textual features from job postings and resumes were vectorized using embedding models, enabling similarity scoring across multiple attributes. Threshold-based ranking filtered candidate matches, while iterative weighting schemes prioritized skill relevance, experience, and qualifications.

"TransparentFit-AI" [24] investigates the role of transparency in AI-supported recruitment systems and examines how disclosure of algorithmic functioning influences human evaluation of person–job compatibility. The framework analyzes external and functional transparency within an experimental resume-screening scenario involving participants with recruitment experience. Findings highlight that improved transparency can reduce disagreement between AI recommendations and recruiter judgement while algorithmic literacy moderates this relationship. The research contributes valuable insight into trust formation and human AI collaboration in hiring environments. Methodological scope remains largely perception-oriented and behavioral, with limited exploration of computational modeling techniques, predictive evaluation metrics, or large-scale recruitment datasets for algorithmic performance validation.

"GraphHire-Net" [25] introduces a recruitment recommendation framework based on graph learning that models relationships among job seekers, employment attributes, and available job opportunities. Candidate profiles and job descriptions are represented as interconnected nodes within a graph structure, allowing relational patterns and contextual associations to guide recommendation generation. The approach demonstrates effectiveness in capturing complex interactions within recruitment ecosystems and improving recommendation relevance. The framework emphasizes representation learning and relational inference within large recruitment datasets. Analytical discussion concentrates mainly on recommendation capability, while interpretability of learned representations, transparency of decision pathways, and deeper evaluation of competency distribution or candidate

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

profile structure receive comparatively limited attention.

Recent advancements in machine learning and intelligent analytical frameworks have demonstrated strong capability in handling complex, high-dimensional data environments, which is essential for modern recruitment analytics. Pattern discovery techniques applied in domains such as cybercrime analysis highlight the importance of extracting structured insights from heterogeneous datasets, a requirement equally relevant in candidate data interpretation and recruitment decision support [26]. Deep learning-based approaches designed for dynamic and adversarial environments emphasize the need for robust analytical mechanisms capable of managing variability and uncertainty within candidate profiles [27]. Emerging real-time data processing paradigms further underline the necessity for efficient and scalable analytical models in handling large volumes of recruitment data [28]. Optimization-driven classification techniques, including particle swarm optimization and genetic algorithm-based methods, demonstrate enhanced feature selection and classification reliability, supporting effective candidate evaluation processes [29], [30]. Adaptive frameworks utilizing fuzzy-based modelling approaches reinforce the importance of handling ambiguity and overlapping characteristics within candidate attributes [31]. Artificial intelligence-driven decision models in business and management contexts highlight the role of intelligent analytics in supporting structured decision-making, which directly aligns with recruitment selection processes [32]. Additionally, adaptive learning perspectives emphasize the significance of contextual understanding in evaluating candidate competencies within diverse recruitment environments [33]. These developments collectively support the need for advanced clustering-based analytical frameworks tailored for intelligent recruitment data analysis and competency-driven candidate classification. In addition, optimization-based intrusion detection frameworks demonstrate the effectiveness of adaptive learning and performance optimization strategies in dynamic environments, which can be extended to recruitment data analysis for improved classification stability and decision reliability [34], [35]. Collectively, these developments emphasize the need for advanced clustering-based analytical frameworks tailored for intelligent recruitment data analysis and competency-driven candidate classification.

Existing recruitment studies primarily focus on recommendation accuracy, transparency analysis, and relational modelling across candidate and job attributes. These approaches provide limited support for competency-oriented grouping within structured recruitment datasets. The dataset used in this study contains heterogeneous candidate attributes, requiring analytical mechanisms capable of capturing density variation and overlapping competency relationships. Such requirements motivate clustering-based analytical design aligned with dataset characteristics.

2.1. Research Gap

Rapid expansion of data-driven recruitment platforms has encouraged the adoption of artificial intelligence for candidate evaluation and job matching. Existing recruitment analytics frameworks mainly concentrate on recommendation mechanisms, transparency evaluation, or behavioural interpretation of human–AI decision interactions. Approaches focusing on algorithmic transparency provide valuable insight into trust formation and recruiter perception; nevertheless, such frameworks emphasize behavioural evaluation rather than computational modelling of candidate competency relationships. Recruitment recommendation models based on relational learning or graph representation primarily target job–candidate matching efficiency, emphasizing recommendation accuracy rather than structured competency grouping within candidate datasets.

Current studies demonstrate limited exploration of clustering-based recruitment analytics capable of identifying competency structures within large candidate pools. Absence of adaptive clustering mechanisms restricts the capability to analyze variations in candidate attribute density and competency distribution. Traditional clustering techniques frequently rely on uniform distance measurements and static attribute weighting strategies, which may not adequately capture the heterogeneous nature of recruitment data. Such limitations reduce the ability to identify meaningful candidate segments that represent diverse skill combinations, experience patterns, and competency characteristics present within recruitment datasets.

Another limitation in existing recruitment analytics research involves insufficient integration of density-aware analysis and flexible membership modelling. Recruitment datasets typically contain overlapping candidate profiles where rigid cluster assignments may not accurately represent competency similarity. Conventional approaches often fail to address such overlapping structures, resulting in limited interpretability of candidate grouping and reduced analytical precision. Lack of adaptive weighting strategies further constrains the capability to represent varying importance of candidate attributes during similarity evaluation. These limitations indicate the necessity for advanced recruitment analytics frameworks capable of integrating adaptive density evaluation, flexible membership modelling, and weighted similarity analysis. Development of such analytical approaches can improve candidate competency interpretation and strengthen classification consistency within recruitment data analysis environments.

The study follows a quantitative experimental design based on recruitment dataset analysis. Candidate records undergo preprocessing, feature transformation, and clustering-based modelling. The proposed framework is evaluated against baseline models using standard classification metrics. Comparative analysis ensures objective assessment of clustering effectiveness and classification reliability.

3. FUZZY-ADAPTIVE DENSITY-WEIGHTED K-MEANS

Fuzzy-Adaptive Density-Weighted K-Means represents a clustering framework integrating fuzzy membership assignment, adaptive density evaluation, and attribute-weighted distance measurement to form competency-oriented clusters, enabling flexible data grouping through iterative centroid refinement and density-sensitive similarity assessment.

3.1. Data Acquisition and Structural Organization of Recruitment Records

Recruitment dataset structuring establishes the foundational representation required for clustering-oriented candidate profiling. Structured representation enables analytical interpretation of applicant characteristics collected through recruitment portals, academic records, technical evaluation platforms, and communication skill assessments. Human-resource evaluation environments normally operate through systematic examination of multiple attributes including qualification level, skill competency, interview evaluation score, and domain knowledge rating. A recruitment dataset can therefore be expressed as a multidimensional matrix containing candidate records across diverse attributes. The structured dataset representation is mathematically defined in Equation (1), where each row corresponds to a candidate record and each column represents a measurable attribute used during recruitment assessment.

$$D = [d_{ij}]_{m\times n} \qquad (1)$$

In Equation (1), D represents the structured recruitment dataset, m denotes the total number of candidate records available for evaluation, n indicates the number of measurable attributes used in recruitment analytics, and d_{ij} refers to the value associated with the i^{th} candidate across the j^{th} attribute. Structured dataset organization allows analytical models to observe relationships among candidate competency indicators in a consistent numerical representation suitable for clustering operations.

Recruitment decision environments rely on evaluation of multiple competency indicators reflecting academic preparation and professional capability. Candidate feature construction transforms raw recruitment information into analytical variables suitable for computational processing. Each candidate profile is represented through an attribute vector describing academic performance, domain knowledge evaluation, communication ability score, and recruitment test outcome. Analytical representation of candidate features is defined in Equation (2), which constructs an attribute vector describing the competency characteristics associated with a candidate record.

$$X_i = (x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}) \qquad (2)$$

In Equation (2), X_i represents the attribute vector describing the i^{th} candidate profile, while x_{ij} denotes the quantitative representation of the j^{th} attribute associated with that candidate. The dimension n reflects the total number of competency indicators incorporated into the recruitment evaluation framework. Analytical construction of feature vectors enables systematic comparison among candidate records and supports the formation of similarity structures within multidimensional recruitment datasets.

Recruitment datasets commonly contain attributes expressed in different numerical scales such as academic grades, interview scores, and communication ratings. Analytical comparison among candidate records requires uniform scaling of attributes to maintain structural consistency within the dataset. Normalization transforms attribute values into a consistent range that prevents dominance of attributes possessing larger numerical

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

magnitude. The normalization operation applied to recruitment attributes is defined in Equation (3), where each attribute value is scaled relative to its observed minimum and maximum range.

$$z_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \qquad (3)$$

In Equation (3), z_{ij} denotes the normalized value of the j^{th} attribute for the i^{th} candidate record, while x_{ij} represents the original attribute value obtained from the recruitment dataset. The quantities $\min(x_j)$ and $\max(x_j)$ correspond to the minimum and maximum observed values of the j^{th} attribute across all candidate records. Normalization ensures balanced representation of attributes and improves the stability of analytical models used in candidate clustering.

Structured recruitment datasets contain diverse candidate profiles distributed across multiple competency attributes. Analytical understanding of this distribution assists recruitment analytics frameworks in recognizing concentration patterns within applicant records. Candidate distribution density across attributes can be expressed through a probabilistic representation that reflects the proportion of candidates exhibiting a particular competency range. The probability distribution describing candidate attribute occurrence is represented in Equation (4).

$$P(x_j) = \frac{f(x_j)}{\sum_{k=1}^{n} f(x_k)} \qquad (4)$$

Within Equation (4), $P(x_j)$ denotes the probability associated with the occurrence of attribute x_j across recruitment records, while $f(x_j)$ represents the observed frequency of that attribute value across candidate profiles. The denominator term represents the cumulative frequency of all considered attributes across the dataset. Distribution analysis supports analytical understanding of candidate competency concentration within recruitment environments.

Recruitment analytics frequently assigns analytical importance to specific attributes that influence candidate selection processes. Academic performance indicators, communication capability, or domain knowledge assessment may contribute differently toward employability evaluation. Structured weighting enables analytical emphasis on attributes that significantly influence recruitment decisions. Attribute weighting for candidate evaluation is represented in Equation (5), where each attribute contributes proportionally to overall candidate evaluation strength.

$$w_j = \frac{s_j}{\sum_{k=1}^{n} s_k} \qquad (5)$$

In Equation (5), w_j denotes the normalized weight associated with the j^{th} attribute, while s_j represents the significance score assigned to that attribute based on recruitment evaluation relevance. The denominator term represents the cumulative significance score across all attributes. Weighted representation improves analytical sensitivity toward attributes considered important within recruitment evaluation environments.

Optimization principles enhance recruitment dataset structuring by ensuring balanced representation of candidate attributes prior to clustering operations. Structured candidate evaluation scores can be represented through a weighted aggregation model that integrates normalized attributes and attribute importance values. Analytical formulation of the candidate evaluation score is defined in Equation (6), which aggregates normalized attribute values through weight coefficients.

$$S_i = \sum_{j=1}^{n} w_j z_{ij} \qquad (6)$$

Within Equation (6), $S_i$represents the aggregated evaluation score associated with the i^{th} candidate profile. The term w_j corresponds to the weight assigned to the j^{th} attribute, while z_{ij} denotes the normalized value of that attribute for the candidate record. Optimization-oriented aggregation improves the interpretability of candidate competency representation within recruitment analytics frameworks.

Analytical preparation of recruitment datasets requires verification of structural completeness before clustering operations commence. Dataset readiness verification ensures balanced representation of candidate records and attribute distributions across the analytical matrix. A dataset readiness indicator can be represented through Equation (7), which evaluates structural completeness relative to attribute coverage within candidate records.

$$R = \frac{1}{m} \sum_{i=1}^{m} \frac{\sum_{j=1}^{n} I(d_{ij})}{n} \qquad (7)$$

In Equation (7), Rrepresents the dataset readiness indicator describing the structural completeness of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

recruitment records. The function $I(d_{ij})$ evaluates the availability of the j^{th} attribute value for the i^{th} candidate record, producing a value of one for available entries and zero for missing entries. The variable $m$represents the total number of candidate profiles and $n$indicates the number of attributes considered during recruitment analysis. Dataset readiness verification supports optimized analytical performance prior to clustering operations.

3.2. Candidate Attribute Identification and Feature Space Definition

Multidimensional candidate feature extraction constructs an analytical representation of recruitment attributes that describe employability characteristics across academic and professional dimensions. Recruitment evaluation environments frequently examine indicators such as academic performance, domain knowledge proficiency, communication skill assessment, technical test performance, and interview evaluation outcomes. Structured representation of these attributes within a multidimensional feature space enables analytical comparison of candidate competencies during clustering analysis. Feature extraction begins through identification of attribute categories that represent measurable employability indicators relevant to recruitment evaluation. The multidimensional feature space for candidate representation can be expressed through Equation (8), which defines the analytical mapping of recruitment attributes into a structured feature environment suitable for computational processing and pattern discovery within candidate datasets.

$$F = \{f_1, f_2, f_3, \dots, f_p\} \qquad (8)$$

In Equation (8), the symbol Frepresents the multidimensional feature space composed of recruitment attributes selected for candidate evaluation. The elements $f_1, f_2, f_3, \dots, f_p$ correspond to individual features representing employability indicators such as academic achievement measures, skill competency evaluations, or communication ratings. The variable p denotes the total number of features extracted from the structured recruitment dataset. Analytical definition of the feature space establishes the dimensional environment required for clustering-based candidate profiling.

Feature extraction involves transformation of candidate records into structured feature representations that preserve relevant competency indicators used in recruitment evaluation. Transformation processes convert normalized dataset attributes obtained from the dataset structuring phase into a feature matrix that represents candidate characteristics within a multidimensional analytical environment. This transformation enables systematic comparison of candidate profiles across multiple competency indicators. Analytical mapping of structured candidate attributes into a feature matrix is expressed in Equation (9), which constructs a representation suitable for multidimensional clustering analysis.

$$\Phi = [\phi_{ij}]_{m \times p} \qquad (9)$$

In Equation (9), Φ denotes the extracted feature matrix representing candidate records across multiple competency attributes. The variable m indicates the number of candidate records present within the recruitment dataset, while p represents the number of extracted features selected for analytical evaluation. The element ϕ_{ij} corresponds to the value associated with the i^{th} candidate across the j^{th} feature dimension. Structural mapping of candidate attributes into a feature matrix enables computational models to interpret employability characteristics through multidimensional analytical representations.

Recruitment evaluation environments prioritize attributes that provide meaningful information regarding candidate suitability for professional roles. Feature extraction therefore incorporates relevance estimation mechanisms that identify attributes contributing significant analytical value toward candidate profiling. Relevance estimation quantifies the contribution of each feature toward the overall representation of candidate competencies. Analytical computation of feature relevance can be expressed through Equation (10), which measures the significance of each feature based on the distribution of feature values across candidate records.

$$\lambda_j = \frac{\sum_{i=1}^{m} (\phi_{ij} - \mu_j)^2}{m} \qquad (10)$$

Within Equation (10), the variable λ_j represents the relevance measure associated with the j^{th} feature extracted from recruitment records. The term ϕ_{ij} corresponds to the feature value of the i^{th} candidate for the j^{th} attribute dimension. The quantity μ_j denotes the average value of the j^{th} feature across all candidate records, while $m$represents the number of candidates considered in the dataset. Feature relevance estimation assists recruitment analytics frameworks in emphasizing attributes that exhibit meaningful variation among candidate profiles.

Optimization-oriented feature extraction ensures that candidate attributes contributing stronger discriminatory information receive greater analytical emphasis during clustering analysis. Recruitment evaluation processes frequently prioritize indicators such as domain knowledge or communication ability depending on organizational recruitment priorities. Feature weighting mechanisms adjust the analytical importance assigned to extracted features. Optimization of feature weights can be represented through Equation (11), which assigns proportional importance to extracted features based on relevance estimation outcomes.

$$\omega_j = \frac{\lambda_j}{\sum_{k=1}^{p} \lambda_k} \tag{11}$$

Within Equation (11), the variable ω_j denotes the optimized weight assigned to the j^{th} feature within the multidimensional candidate representation. The value λ_j corresponds to the relevance score computed for that feature, while the denominator term represents the cumulative relevance scores across all extracted features. Optimization-based weighting enhances analytical sensitivity toward attributes exhibiting stronger variation across candidate profiles and improves the quality of clustering outcomes.

Feature extraction culminates through the construction of an optimized competency representation describing each candidate profile. Recruitment analytics frameworks combine extracted feature values and optimized feature weights to represent candidate competencies within the multidimensional analytical environment used for clustering operations. Analytical representation of candidate competency strength can be expressed through Equation (12), which integrates extracted features with optimized weighting coefficients.

$$C_i = \sum_{j=1}^{p} \omega_j \, \phi_{ij} \tag{12}$$

In Equation (12), the variable C_i denotes the competency representation score associated with the i^{th} candidate record. The coefficient ω_j represents the optimized weight assigned to the j^{th} feature dimension, while ϕ_{ij} corresponds to the feature value for that candidate across the selected attribute dimension. The summation aggregates weighted features to represent the competency strength of each candidate profile within recruitment analytics frameworks.

Multidimensional candidate feature extraction requires analytical examination of relationships among extracted attributes. Recruitment datasets frequently exhibit interactions among competency indicators such as academic preparation and domain knowledge evaluation. Analytical modeling of such relationships enhances interpretation of candidate profiles within clustering environments. Interaction strength among extracted features can be represented through Equation (13), which measures pairwise interaction magnitude between features within the multidimensional candidate representation.

$$I_{jk} = \frac{1}{m} \sum_{i=1}^{m} \mid \phi_{ij} - \phi_{ik} \mid \tag{13}$$

In Equation (13), the variable I_{jk} represents the interaction magnitude between the j^{th} and k^{th} features extracted from recruitment records. The values ϕ_{ij} and ϕ_{ik} denote feature values associated with the i^{th} candidate across the respective feature dimensions. The variable $m$represents the number of candidate records used in the analysis. Interaction analysis assists recruitment analytics frameworks in understanding structural relationships present among competency indicators.

Robust feature extraction requires evaluation of feature stability to ensure consistent representation of candidate attributes within clustering environments. Stability analysis examines the distribution of feature values across candidate records and ensures that extracted attributes contribute meaningful structural information to clustering operations. Analytical formulation of feature stability can be represented through Equation (14), which evaluates dispersion patterns within extracted features.

$$\psi_j = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (\phi_{ij} - \bar{\phi}_j)^2} \tag{14}$$

Within Equation (14), the variable ψ_j denotes the stability indicator associated with the j^{th} extracted feature. The term $\bar{\phi}_j$ represents the mean value of that feature across candidate records, while ϕ_{ij} corresponds to the value associated with the i^{th} candidate. The variable m denotes the number of candidate profiles present in the recruitment dataset. Stability evaluation supports optimization of feature representation and enhances reliability of clustering-based candidate profiling.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.3. Analytical Necessity of Attribute Standardization in Recruitment Evaluation

Recruitment datasets contain heterogeneous attributes describing candidate competencies across academic and professional dimensions. Feature extraction processes produce multidimensional representations in which attributes such as academic grades, interview scores, communication evaluations, and technical assessment results appear in different numerical ranges. Direct comparison among such attributes may distort clustering behavior, since attributes with large numerical ranges may dominate similarity computation. Attribute standardization resolves this analytical imbalance through transformation of feature values into comparable scales. A standardized representation ensures that each competency indicator contributes proportionally during clustering analysis. Mathematical representation of standardized feature transformation is expressed in Equation (15), which converts extracted feature values into standardized form based on statistical characteristics of the dataset.

$$s_{ij} = \frac{\phi_{ij} - \mu_j}{\sigma_j} \qquad (15)$$

In Equation (15), the term s_{ij} represents the standardized value of the j^{th} feature associated with the i^{th} candidate record. The symbol ϕ_{ij} denotes the extracted feature value obtained during the multidimensional feature extraction stage. The quantity μ_j represents the mean value of the j^{th} feature across all candidate records, while σ_j indicates the standard deviation associated with that feature. Standardization allows analytical comparison of attributes within a uniform statistical framework.

Standardization requires estimation of central tendency within each feature dimension to determine deviation of candidate attributes from average competency levels. Recruitment analytics environments interpret mean values as indicators of typical performance levels across applicant populations. Mean estimation therefore assists clustering mechanisms in understanding relative candidate strengths across competency indicators. Statistical computation of feature mean is represented through Equation (16), which aggregates feature values across all candidate records.

$$\mu_j = \frac{1}{m}\sum_{i=1}^{m} \phi_{ij} \qquad (16)$$

Within Equation (16), the variable μ_j denotes the mean value of the j^{th} extracted feature across candidate records. The quantity m represents the total number of candidate profiles contained in the recruitment dataset. The element ϕ_{ij} corresponds to the value associated with the i^{th} candidate across the j^{th} feature dimension. Mean estimation provides a reference point for measuring deviation patterns in candidate attributes during standardization.

Candidate datasets frequently contain variation in competency levels across attributes such as academic achievements or interview performance outcomes. Variance analysis measures the spread of feature values across candidate records and supports accurate scaling of attributes within clustering environments. Statistical dispersion of extracted features is represented through Equation (17), which evaluates the magnitude of variation present within each feature dimension.

$$v_j = \frac{1}{m}\sum_{i=1}^{m} (\phi_{ij} - \mu_j)^2 \qquad (17)$$

In Equation (17), the symbol v_j denotes the variance associated with the j^{th} extracted feature. The quantity ϕ_{ij} represents the feature value for the i^{th} candidate, while μ_j corresponds to the mean value of the feature across candidate records. The variable $m$indicates the number of candidate profiles present in the dataset. Variance estimation assists standardization mechanisms in identifying dispersion levels within recruitment attributes.

Recruitment analytics environments benefit from feature scaling approaches that preserve relative relationships among candidate attributes while ensuring uniform value ranges across feature dimensions. Range-based scaling transforms standardized attributes into bounded numerical intervals that support stable clustering behavior. Feature scaling through range transformation is expressed in Equation (18), which converts standardized feature values into a normalized scale.

$$r_{ij} = \frac{s_{ij} - \alpha_j}{\beta_j - \alpha_j} \qquad (18)$$

In Equation (18), the variable r_{ij} represents the scaled value of the j^{th} feature associated with the i^{th} candidate record. The symbol s_{ij} corresponds to the standardized feature value obtained through Equation (15). The parameters α_j and β_j represent the minimum and maximum standardized values observed within the j^{th} feature dimension. Range

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

scaling ensures balanced representation of candidate attributes within clustering operations.

Recruitment evaluation processes often emphasize particular competency indicators considered critical for candidate selection. Optimization-oriented feature scaling integrates attribute importance into the scaling process by adjusting scaled values using feature weights obtained during feature extraction. Weighted scaling improves representation of attributes that significantly influence recruitment decisions. Analytical formulation of optimized scaling is represented through Equation (19).

$$\theta_{ij} = \omega_j \times r_{ij} \qquad (19)$$

Within Equation (19), the variable θ_{ij} represents the optimized scaled value associated with the j^{th} feature for the i^{th} candidate record. The term r_{ij} corresponds to the scaled feature value obtained through range normalization. The coefficient ω_j represents the optimized feature weight calculated during the multidimensional feature extraction stage. Optimization-based scaling ensures that influential attributes maintain appropriate analytical significance within clustering analysis.
Reliable clustering requires verification of standardized feature distributions to ensure balanced analytical representation of candidate attributes. Consistency verification evaluates dispersion patterns within scaled features and confirms that feature distributions remain stable across candidate records. A consistency indicator can be formulated through Equation (20), which measures average magnitude of scaled features across the dataset.

$$\eta_j = \frac{1}{m}\sum_{i=1}^{m} |\,\theta_{ij}\,| \qquad (20)$$

In Equation (20), the symbol η_j denotes the consistency indicator associated with the j^{th} feature dimension. The value θ_{ij} represents the optimized scaled feature value corresponding to the i^{th} candidate record. The variable $m$indicates the total number of candidate profiles contained within the recruitment dataset. Consistency verification ensures stable feature representation prior to clustering operations.

Completion of attribute standardization and feature scaling leads to construction of a scaled feature matrix representing candidate attributes within a uniform analytical space. This matrix provides the optimized feature representation used for subsequent clustering analysis within the FADW-KM framework. The scaled feature matrix is expressed through Equation (21).

$$\Omega = [\theta_{ij}]_{m\times p} \qquad (21)$$

Within Equation (21), the symbol Ω represents the scaled candidate feature matrix prepared for clustering operations. The variable $m$denotes the number of candidate records within the recruitment dataset, while p represents the number of extracted features. The element θ_{ij} corresponds to the optimized scaled value of the j^{th} feature associated with the i^{th} candidate profile. This representation enables stable and balanced clustering of candidate competencies.

3.4. Candidate Competency Space Preparation for Cluster Initialization

Adaptive cluster center initialization begins through interpretation of the scaled feature matrix obtained from the attribute standardization phase. Recruitment analytics environments examine candidate competencies distributed across multiple attributes representing academic preparation, communication skill level, and technical assessment outcomes. Analytical preparation of this competency space ensures that clustering operations begin from representative positions within the multidimensional data structure. A candidate competency representation for initialization can be expressed through Equation (22), which defines the overall competency magnitude associated with each candidate record within the scaled feature matrix.

$$\kappa_i = \sum_{j=1}^{p} (\,\theta_{ij})^2 \qquad (22)$$

In Equation (22), the variable κ_i denotes the competency magnitude associated with the i^{th} candidate record within the scaled feature representation. The symbol θ_{ij} represents the optimized scaled feature value associated with the j^{th} attribute of that candidate profile. The variable p indicates the number of extracted attributes used in recruitment evaluation.

Recruitment datasets frequently exhibit diverse candidate profiles across competency attributes. Analytical evaluation of candidate distribution assists initialization mechanisms in identifying suitable starting regions for clustering. Distribution estimation quantifies the average competency value across candidate records, allowing identification of central tendencies within the dataset. Analytical estimation of candidate competency distribution is expressed in Equation (23).

$$\delta = \frac{1}{m}\sum_{i=1}^{m}\kappa_i \qquad (23)$$

In Equation (23), the symbol δ represents the average competency magnitude across candidate records. The variable m denotes the number of candidate profiles contained within the recruitment dataset. The quantity κ_i corresponds to the competency magnitude calculated in Equation (22). Distribution estimation provides an analytical basis for determining suitable cluster initialization regions.

Cluster center initialization requires selection of candidate profiles that represent distinct competency characteristics within the dataset. Recruitment evaluation concepts mirror analytical selection of representative profiles that reflect varying competency strengths across applicant populations. Analytical identification of reference candidates can be represented through Equation (24), which selects candidate records exhibiting competency magnitude exceeding the dataset distribution indicator.

$$\rho_i = \begin{cases} 1, & \kappa_i > \delta \\ 0, & \text{otherwise} \end{cases} \qquad (24)$$

In Equation (24), the indicator variable ρ_i determines candidate eligibility for cluster center initialization. A value of one indicates that the competency magnitude of the candidate exceeds the dataset average. A value of zero indicates lower competency magnitude relative to the dataset distribution level.

Adaptive cluster center generation constructs initial cluster centers using selected candidate references derived from competency distribution analysis. Analytical formulation of initial cluster centers integrates scaled feature values associated with selected candidate profiles. The initial cluster center representation is defined in Equation (25).

$$c_{kj}^{(0)} = \frac{\sum_{i=1}^{m}\rho_i\,\theta_{ij}}{\sum_{i=1}^{m}\rho_i} \qquad (25)$$

Within Equation (25), the variable $c_{kj}^{(0)}$ denotes the initial value of the j^{th} feature component for the k^{th} cluster center. The term ρ_i represents the reference indicator defined in Equation (24), while θ_{ij} corresponds to the scaled feature value for the candidate record. The denominator represents the total number of candidate references selected for initialization.

Effective clustering requires sufficient separation among initialized cluster centers to avoid redundant grouping patterns. Recruitment analytics frameworks benefit from analytical separation measures that ensure diversity among competency groups. Cluster separation distance can be defined through Equation (26).

$$\Gamma_{kl} = \sqrt{\sum_{j=1}^{p}(c_{kj}^{(0)} - c_{lj}^{(0)})^2} \qquad (26)$$

In Equation (26), the symbol Γ_{kl}represents the separation distance between the k^{th} and l^{th} initialized cluster centers. The quantities $c_{kj}^{(0)}$ and $c_{lj}^{(0)}$ denote the feature values associated with the corresponding cluster centers across the j^{th} attribute dimension.

Adaptive initialization incorporates an optimization condition that ensures cluster centers maintain sufficient separation while representing candidate competency patterns accurately. Analytical evaluation of initialization quality can be expressed through Equation (27).

$$Q = \sum_{k=1}^{K}\sum_{l=1}^{K}\Gamma_{kl} \qquad (27)$$

In Equation (27), the symbol Q represents the initialization quality score measuring cumulative separation among cluster centers. The parameter K indicates the total number of clusters intended for formation within the recruitment dataset. Larger values of Q correspond to improved separation among candidate competency groups.

Optimization-driven initialization may refine cluster centers to ensure balanced representation of candidate competency structures. Adjustment of cluster centers based on competency magnitude can be expressed through Equation (28).

$$\tilde{c}_{kj} = c_{kj}^{(0)} + \frac{\delta}{1+|\,c_{kj}^{(0)}\,|} \qquad (28)$$

Within Equation (28), the variable $\tilde{c}_{kj}$ represents the adjusted value of the j^{th} feature component for the k^{th} cluster center. The term $c_{kj}^{(0)}$ denotes the initial cluster center value obtained through Equation (25), while δ represents the dataset competency distribution level.

Completion of initialization produces a matrix containing optimized cluster center representations used during clustering iterations. The cluster initialization matrix can be represented through Equation (29).

$$C^{(0)} = [\tilde{c}_{kj}]_{K\times p} \qquad (29)$$

In Equation (29), the symbol $C^{(0)}$ represents the initialized cluster center matrix. The parameter K denotes the number of clusters, while p indicates the number of extracted candidate attributes. The element $\tilde{c}_{kj}$ corresponds to the adjusted value of the j^{th} feature component for the k^{th} cluster center.

3.5. Analytical Significance of Local Candidate Density in Recruitment Evaluation

Local density estimation examines the concentration of candidate profiles within the multidimensional competency space constructed through earlier methodological stages. Recruitment analytics environments frequently contain groups of applicants exhibiting comparable academic preparation, technical expertise, and communication capability. Human-resource decision processes often recognize such concentration patterns during evaluation of candidate suitability for particular roles. Analytical identification of candidate concentration assists clustering mechanisms in recognizing meaningful competency regions within the dataset. A mathematical representation of local candidate density begins through measurement of neighborhood proximity around each candidate profile. The proximity-based neighborhood indicator is expressed through Equation (30), which defines the interaction condition between candidate records within the scaled competency space.

$$\zeta_{il} = \begin{cases} 1, & \text{if } \Delta_{il} \le \tau \\ 0, & \text{otherwise} \end{cases} \qquad (30)$$

In Equation (30), the symbol ζ_{il} represents the neighborhood interaction indicator between the i^{th} candidate and the l^{th} candidate. The quantity Δ_{il} denotes the proximity distance between the candidate records, while the parameter τ represents the neighborhood influence threshold defining the maximum distance within which interaction is considered meaningful for density estimation.

Density estimation requires analytical measurement of proximity among candidate profiles across the multidimensional competency attributes generated through earlier processing stages. Proximity measurement evaluates similarity between candidate records based on scaled feature representation. Analytical computation of candidate proximity distance is expressed through Equation (31), which measures separation among candidate records within the competency space.

$$\Delta_{il} = \sqrt{\sum_{j=1}^{p} (\theta_{ij} - \theta_{lj})^2} \qquad (31)$$

In Equation (31), the symbol Δ_{il} represents the proximity distance between candidate records indexed by i and l. The quantity θ_{ij} corresponds to the scaled feature value associated with the j^{th} attribute of the i^{th} candidate record, while θ_{lj} denotes the corresponding attribute value for the l^{th} candidate. The variable $p$indicates the number of competency attributes included within the analytical representation.

Local density estimation evaluates the concentration of candidate records surrounding each candidate profile within the competency space. Higher density values correspond to candidate regions where many applicants exhibit comparable competency characteristics. Such concentration patterns resemble recruitment situations where multiple applicants demonstrate similar qualification levels and skill competencies. Analytical computation of candidate density is expressed through Equation (32), which aggregates neighborhood interactions across candidate records.

$$\lambda_i = \sum_{l=1}^{m} \zeta_{il} \qquad (32)$$

In Equation (32), the variable λ_i represents the local density associated with the i^{th} candidate profile. The indicator ζ_{il} represents the neighborhood interaction condition defined in Equation (30). The variable m denotes the number of candidate records contained within the recruitment dataset. Local density values provide analytical insight regarding the concentration of competency patterns among candidate groups.

Recruitment analytics environments benefit from analytical measures that estimate the influence of candidate profiles located within high-density regions. Candidate records positioned within dense competency clusters frequently represent common qualification patterns observed within applicant populations. Analytical representation of density-based influence can be expressed through Equation (33), which computes influence magnitude relative to the density distribution across candidate profiles.

$$\psi_i = \frac{\lambda_i}{\sum_{r=1}^{m} \lambda_r} \quad (33)$$

In Equation (33), the symbol ψ_i denotes the influence factor associated with the i^{th} candidate profile. The quantity λ_i corresponds to the local density value computed in Equation (32), while the denominator represents the cumulative density across all candidate records. The variable m indicates the number of candidate profiles within the dataset.

Density estimation contributes to clustering performance through adaptive weighting of candidate records based on local concentration patterns. Optimization-oriented clustering frameworks assign greater analytical importance to candidate records situated within dense regions of the competency space. Such weighting strategies improve cluster formation stability by emphasizing representative candidate groups. Analytical representation of density-based weighting is expressed through Equation (34).

$$\omega_i^d = \frac{\psi_i}{\max(\psi)} \quad (34)$$

In Equation (34), the variable ω_i^d represents the density-based weight assigned to the i^{th} candidate record. The term ψ_i corresponds to the influence factor computed in Equation (33), while $\max(\psi)$ denotes the maximum influence value observed across candidate profiles. Density-based weighting supports optimization of clustering processes by strengthening representation of concentrated competency patterns.

Reliable density estimation requires evaluation of stability across candidate records to ensure balanced representation of competency concentration patterns. Stability analysis examines variation in density weights across candidate profiles and assists clustering algorithms in maintaining stable grouping behavior. Analytical formulation of density stability can be expressed through Equation (35), which measures dispersion within density weights.

$$\chi = \frac{1}{m} \sum_{i=1}^{m} (\omega_i^d - \bar{\omega}^d)^2 \quad (35)$$

In Equation (35), the symbol χ represents the stability indicator describing dispersion of density weights across candidate profiles. The quantity ω_i^d denotes the density weight associated with the i^{th} candidate record. The variable $\bar{\omega}^d$ represents the average density weight across all candidate records, while m denotes the number of candidate profiles present within the recruitment dataset. Stability evaluation ensures consistent representation of candidate concentration patterns within clustering analysis.

Completion of density estimation results in a structured representation containing density information associated with each candidate record. This representation becomes an essential component for subsequent density-weighted clustering operations within the FADW-KM framework. Analytical construction of the density representation matrix is expressed through Equation (36).

$$D^{\rho} = [\omega_i^d]_{m\times 1} \quad (36)$$

In Equation (36), the symbol D^{ρ} represents the density vector describing the density weight associated with each candidate record. The variable m indicates the number of candidate profiles within the recruitment dataset. The element ω_i^d corresponds to the optimized density weight associated with the i^{th} candidate record.

3.6. Density Influence Interpretation for Candidate Attributes

Density weight computation expands the local density estimation phase through analytical evaluation of attribute-level influence across candidate profiles. Recruitment analytics environments frequently examine how clusters of candidates exhibit concentration across specific competency indicators such as academic achievement, domain expertise, and communication capability. Attribute-level density evaluation supports identification of competency indicators that contribute strongly to concentrated candidate regions. Analytical representation of attribute density influence is formulated through Equation (37), which measures aggregated density contribution of each attribute across candidate records within the competency space.

$$\Lambda_j = \sum_{i=1}^{m} \omega_i^d \cdot \theta_{ij} \quad (37)$$

In Equation (37), the symbol Λ_j represents the density influence associated with the j^{th} candidate attribute. The value ω_i^d denotes the density weight assigned to the i^{th} candidate record during the previous density estimation phase. The variable θ_{ij} corresponds to the scaled attribute value for the j^{th} feature associated with the i^{th} candidate. The variable $m$represents the number of candidate profiles contained within the dataset.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Recruitment evaluation frameworks frequently assess distribution patterns across competency attributes to identify indicators representing strong clustering behavior among candidate profiles. Attribute density distribution analysis examines how density contributions are distributed across all attributes in the dataset. Analytical computation of normalized density distribution for each attribute is expressed through Equation (38), which evaluates the proportional density contribution of each competency indicator.

$$\Pi_j = \frac{\Lambda_j}{\sum_{t=1}^{p} \Lambda_t} \qquad (38)$$

In Equation (38), the symbol Π_j represents the normalized density proportion associated with the j^{th} candidate attribute. The variable Λ_j corresponds to the density influence computed through Equation (37), while the denominator represents cumulative density influence across all attributes contained within the multidimensional competency representation. The parameter $p$denotes the total number of attributes considered during clustering analysis.

Density distribution estimation enables optimization of attribute weighting by emphasizing competency indicators contributing strongly to candidate concentration patterns. Recruitment decision environments often prioritize attributes that demonstrate meaningful variation among applicant groups. Density-based optimization integrates attribute distribution patterns with clustering operations by generating density weights for each competency indicator. Analytical formulation of optimized density weight assignment is represented through Equation (39).

$$\Omega_j^d = \frac{\Pi_j}{\max(\Pi)} \qquad (39)$$

In Equation (39), the variable Ω_j^d represents the optimized density weight associated with the j^{th} candidate attribute. The value Π_j denotes the normalized density proportion obtained through Equation (38), while $\max(\Pi)$ represents the maximum density proportion observed across the set of candidate attributes. Density weight optimization strengthens analytical emphasis on attributes exhibiting strong candidate concentration patterns.

Density weights computed for candidate attributes contribute directly to clustering operations through weighted integration of attribute values. Weighted integration ensures that attributes representing concentrated competency patterns exert greater influence during cluster formation. Analytical formulation of weighted attribute density integration is expressed through Equation (40), which combines density weights with scaled feature values for each candidate record.

$$\Gamma_{ij} = \Omega_j^d \times \theta_{ij} \qquad (40)$$

Within Equation (40), the symbol Γ_{ij} represents the density-weighted attribute value associated with the j^{th} feature of the i^{th} candidate record. The coefficient Ω_j^d denotes the density weight assigned to the attribute, while θ_{ij} corresponds to the scaled attribute value obtained during the feature scaling stage. This integration process strengthens representation of attributes demonstrating strong density influence.

Reliable density weighting requires evaluation of stability across candidate attributes to ensure balanced representation within clustering operations. Stability analysis examines variation among density-weighted attributes across candidate records and supports optimization of density weighting strategies. Analytical representation of attribute density stability is expressed through Equation (41), which measures dispersion within density-weighted attributes.

$$\Phi_j^d = \frac{1}{m}\sum_{i=1}^{m} (\Gamma_{ij} - \bar{\Gamma}_j)^2 \qquad (41)$$

In Equation (41), the symbol Φ_j^d represents the stability indicator associated with the j^{th} density-weighted attribute. The value Γ_{ij} corresponds to the density-weighted attribute value associated with the i^{th} candidate record. The quantity $\bar{\Gamma}_j$ represents the mean density-weighted value for that attribute across candidate records. The parameter $m$denotes the number of candidate profiles present within the recruitment dataset.

Completion of density weight computation results in construction of a density-weighted feature representation that integrates attribute scaling, density influence, and optimization-based weighting mechanisms. This representation forms the analytical foundation for fuzzy membership allocation and cluster optimization stages within the FADW-KM framework. Mathematical representation of the density-weighted feature matrix is expressed through Equation (42).

$$\Psi = [\Gamma_{ij}]_{m \times p} \qquad (42)$$

In Equation (42), the symbol Ψ denotes the density-weighted feature matrix used for clustering analysis. The parameter m represents the number of candidate records within the recruitment dataset, while p indicates the number of competency attributes included in the multidimensional feature representation. The element Γ_{ij} corresponds to the density-weighted attribute value associated with the i^{th} candidate across the j^{th} feature dimension.

3.7. Analytical Representation of Candidate–Cluster Proximity

Weighted distance similarity computation evaluates proximity between candidate profiles and initialized cluster centers using the density-weighted feature matrix obtained in the previous stage. Recruitment analytics environments frequently involve comparison of candidate competencies with representative profiles reflecting particular skill groups. Analytical evaluation of similarity enables identification of candidate records that exhibit competency characteristics aligned with specific cluster centers. Mathematical formulation of the basic proximity relationship between candidate attributes and cluster centers is expressed through Equation (43).

$$\Delta_{ik}^{w} = \sqrt{\sum_{j=1}^{p} (\Gamma_{ij} - \tilde{c}_{kj})^2} \qquad (43)$$

In Equation (43), the symbol Δ_{ik}^{w} denotes the weighted proximity distance between the i^{th} candidate record and the k^{th} cluster center. The quantity Γ_{ij} represents the density-weighted attribute value associated with the j^{th} competency indicator of the candidate profile, while $\tilde{c}_{kj}$ denotes the value of the j^{th} attribute associated with the initialized cluster center. The variable $p$corresponds to the total number of attributes within the candidate representation.

Recruitment evaluation procedures frequently assign varying levels of importance to competency attributes such as domain knowledge or communication capability. Weighted similarity computation integrates attribute importance to ensure that influential attributes contribute proportionally during proximity evaluation. Analytical incorporation of attribute importance within similarity measurement is expressed through Equation (44), which modifies distance computation through attribute-level weighting.

$$\Delta_{ik}^{a} = \sqrt{\sum_{j=1}^{p} \Omega_j^d \, (\Gamma_{ij} - \tilde{c}_{kj})^2} \qquad (44)$$

In Equation (44), the symbol Δ_{ik}^{a} represents the attribute-weighted distance between the i^{th} candidate profile and the k^{th} cluster center. The coefficient Ω_j^d denotes the density-based weight assigned to the j^{th} competency attribute. The variables Γ_{ij} and $\tilde{c}_{kj}$ correspond to density-weighted candidate attributes and cluster center attributes respectively.

Similarity evaluation benefits from integration of candidate density information obtained during earlier analytical stages. Candidate profiles located within dense competency regions often represent typical skill combinations observed within recruitment datasets. Density influence adjustment strengthens similarity computation for candidate profiles located within such regions. Analytical formulation of density-adjusted distance computation is represented through Equation (45).

$$\Delta_{ik}^{d} = \omega_i^d \times \Delta_{ik}^{a} \qquad (45)$$

Within Equation (45), the variable Δ_{ik}^{d} represents the density-adjusted similarity distance between candidate record i and cluster center k. The coefficient ω_i^d denotes the density weight assigned to the candidate record during local density estimation. The quantity Δ_{ik}^{a} corresponds to the attribute-weighted distance obtained in Equation (44).

Distance values obtained through weighted computation require transformation into similarity measures suitable for clustering evaluation. Similarity representation converts proximity measurements into bounded values that reflect closeness between candidate competencies and cluster centers. Analytical transformation of weighted distances into similarity scores is expressed through Equation (46).

$$S_{ik} = \frac{1}{1 + \Delta_{ik}^{d}} \qquad (46)$$

In Equation (46), the variable S_{ik} denotes the similarity score associated with the i^{th} candidate record relative to the k^{th} cluster center. The term Δ_{ik}^{d} corresponds to the density-adjusted distance computed in Equation (45). Higher similarity values indicate stronger alignment between candidate competencies and cluster characteristics.

Clustering operations require normalized similarity measures across cluster centers to ensure balanced candidate evaluation. Normalization ensures that similarity scores represent relative proximity among cluster centers rather than absolute distance magnitudes. Analytical formulation of similarity normalization is represented through Equation (47).

$$\hat{S}_{ik} = \frac{S_{ik}}{\sum_{k=1}^{K} S_{ik}} \quad (47)$$

In Equation (47), the symbol $\hat{S}_{ik}$ represents the normalized similarity score associated with the i^{th} candidate record and the k^{th} cluster center. The parameter K denotes the number of clusters present within the clustering model. The denominator represents the cumulative similarity scores for the candidate record across all cluster centers. Optimization strategies strengthen clustering performance through reinforcement of strong similarity relationships between candidate profiles and cluster centers. Reinforcement mechanisms emphasize high similarity scores while reducing influence of weak similarity relationships. Analytical representation of similarity reinforcement is expressed through Equation (48).

$$R_{ik} = (\hat{S}_{ik})^{\gamma} \quad (48)$$

In Equation (48), the variable R_{ik} denotes the reinforced similarity score between candidate record $i$and cluster center k. The quantity $\hat{S}_{ik}$ represents the normalized similarity score defined in Equation (47). The exponent parameter γrepresents the reinforcement coefficient controlling the strength of similarity emphasis.

Reliable clustering operations require evaluation of similarity stability across candidate records to ensure consistent representation of proximity relationships. Stability analysis examines dispersion within reinforced similarity scores and ensures balanced similarity distribution among candidate profiles. Analytical representation of similarity stability is expressed through Equation (49).

$$\Xi = \frac{1}{mK} \sum_{i=1}^{m} \sum_{k=1}^{K} (R_{ik} - \bar{R})^2 \quad (49)$$

In Equation (49), the symbol Ξ represents the similarity stability indicator across candidate records. The variable R_{ik} denotes the reinforced similarity value associated with candidate record i and cluster center k. The quantity $\bar{R}$represents the average reinforced similarity across the candidate–cluster similarity matrix.

Completion of weighted distance similarity computation produces a similarity matrix describing proximity relationships between candidate records and cluster centers. This matrix serves as the analytical foundation for fuzzy membership allocation in subsequent clustering stages. Mathematical representation of the weighted similarity matrix is expressed through Equation (50).

$$\Theta = [R_{ik}]_{m \times K} \quad (50)$$

In Equation (50), the symbol Θ represents the reinforced similarity matrix used for clustering operations. The parameter m denotes the number of candidate records within the recruitment dataset, while K indicates the number of clusters defined in the clustering framework. The element R_{ik} corresponds to the reinforced similarity value associated with candidate record i relative to cluster center k.

3.8. Fuzzy Representation of Candidate–Cluster Association

Fuzzy membership degree assignment transforms reinforced similarity relationships obtained during weighted similarity computation into probabilistic association values describing candidate alignment with cluster centers. Recruitment analytics environments frequently observe candidates exhibiting competency characteristics suitable for multiple skill groups rather than a single rigid category. Analytical modeling of such overlapping competency patterns requires flexible membership representation. Fuzzy membership computation converts similarity relationships into normalized association measures describing candidate affiliation with each cluster. Analytical representation of preliminary membership assignment is defined in Equation (51).

$$\mu_{ik} = \frac{R_{ik}}{\sum_{k=1}^{K} R_{ik}} \quad (51)$$

In Equation (51), the symbol μ_{ik} represents the preliminary fuzzy membership value describing the association of the i^{th} candidate record with the k^{th} cluster center. The variable R_{ik} denotes the reinforced similarity value computed during weighted distance similarity analysis. The parameter K represents the number of clusters within the clustering framework. Normalization across clusters ensures balanced membership representation for each candidate profile.

Fuzzy clustering frameworks require enforcement of membership constraints to maintain consistent interpretation of candidate association values. Each candidate profile must maintain a total membership contribution equal to unity across all clusters, ensuring that the candidate competency representation remains proportionally distributed among cluster centers. Analytical enforcement of the fuzzy constraint condition is expressed through Equation (52).

$$\sum_{k=1}^{K} \mu_{ik} = 1 \qquad (52)$$

Within Equation (52), the summation of fuzzy membership values across all clusters equals unity for each candidate record. The variable μ_{ik} represents the membership value associated with the i^{th} candidate and the k^{th} cluster center. This constraint preserves interpretability of candidate affiliation across competency groups.

Recruitment analytics often benefits from reinforcement of strong candidate–cluster associations that reflect meaningful competency alignment. Optimization-oriented membership reinforcement emphasizes clusters exhibiting strong similarity relationships while preserving fuzzy representation of overlapping competencies. Analytical formulation of reinforced membership assignment is expressed through Equation (53).

$$\tilde{\mu}_{ik} = \frac{(\mu_{ik})^{\beta}}{\sum_{k=1}^{K}(\mu_{ik})^{\beta}} \qquad (53)$$

In Equation (53), the symbol $\tilde{\mu}_{ik}$ denotes the reinforced fuzzy membership value for the i^{th} candidate record relative to the k^{th} cluster center. The variable μ_{ik} corresponds to the preliminary membership value computed through Equation (51). The parameter β represents the membership reinforcement coefficient controlling the strength of fuzzy emphasis within clustering analysis.

Reliable fuzzy clustering requires examination of membership stability across candidate records to ensure consistent representation of candidate affiliation patterns. Stability evaluation measures dispersion among reinforced membership values across cluster centers. Analytical representation of membership stability is defined through Equation (54).

$$\Upsilon = \frac{1}{mK} \sum_{i=1}^{m} \sum_{k=1}^{K} (\tilde{\mu}_{ik} - \bar{\mu})^2 \qquad (54)$$

In Equation (54), the symbol Υrepresents the stability indicator describing variation among reinforced membership values. The quantity $\tilde{\mu}_{ik}$ denotes the reinforced fuzzy membership value for candidate i and cluster k, while $\bar{\mu}$ represents the average membership value across all candidate–cluster associations. The parameter m indicates the number of candidate profiles present within the recruitment dataset.

Completion of fuzzy membership assignment results in formation of a structured matrix describing candidate affiliation across all cluster centers. This matrix forms the analytical basis for cluster center optimization during iterative clustering stages. Mathematical representation of the fuzzy membership matrix is defined through Equation (55).

$$M = [\tilde{\mu}_{ik}]_{m \times K} \qquad (55)$$

In Equation (55), the symbol Mrepresents the fuzzy membership matrix used within the clustering framework. The variable m denotes the number of candidate records, while K corresponds to the number of clusters within the clustering model. The element $\tilde{\mu}_{ik}$ describes the reinforced membership value representing the degree of association between candidate record i and cluster center k.

3.9. Density-Adaptive Refinement of Cluster Centers

Density-adaptive cluster center optimization refines cluster representatives through integration of fuzzy membership relationships and candidate density influence obtained in previous stages. Recruitment analytics environments often require identification of representative candidate competency patterns that accurately reflect dominant skill distributions within applicant populations. Adaptive refinement adjusts cluster centers according to candidate records exhibiting strong membership association and significant density influence. Analytical formulation of the optimized cluster center position is expressed through Equation (56).

$$c_{kj}^{*} = \frac{\sum_{i=1}^{m}(\tilde{\mu}_{ik})^{\beta} \cdot \omega_i^d \cdot \Gamma_{ij}}{\sum_{i=1}^{m}(\tilde{\mu}_{ik})^{\beta} \cdot \omega_i^d} \qquad (56)$$

In Equation (56), the symbol c_{kj}^{*} represents the optimized value of the j^{th} attribute associated with the k^{th}cluster center. The variable $\tilde{\mu}_{ik}$ denotes the reinforced fuzzy membership value assigned to the i^{th}candidate for the k^{th} cluster. The exponent parameter β controls membership influence during optimization. The coefficient ω_i^d corresponds to the density weight assigned to the candidate record during local density estimation. The value Γ_{ij} represents the density-weighted attribute value associated with the j^{th} feature of the candidate record. The parameter m denotes the total number of candidate profiles within the dataset.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Cluster center optimization requires evaluation of clustering quality through an objective function measuring compactness of candidate groups around optimized cluster centers. Recruitment analytics environments conceptually resemble evaluation procedures conducted by human-resource analysts who attempt to identify representative competency profiles among applicant populations. Analytical representation of clustering compactness is expressed through Equation (57).

$$J = \sum_{i=1}^{m} \sum_{k=1}^{K} (\tilde{\mu}_{ik})^{\beta} \cdot \omega_i^d \cdot \| \Gamma_i - C_k^* \|^2 \qquad (57)$$

In Equation (57), the symbol Jrepresents the clustering objective function measuring overall compactness of candidate records around optimized cluster centers. The vector Γ_i denotes the density-weighted feature vector representing the i^{th} candidate profile, while C_k^* represents the optimized feature vector associated with the k^{th} cluster center. The term $(\tilde{\mu}_{ik})^{\beta}$ reflects reinforced fuzzy membership influence, and ω_i^d denotes density influence associated with the candidate profile. The parameter K represents the number of clusters.

Optimization of cluster centers continues iteratively until minimal variation occurs between successive cluster center updates. Convergence evaluation ensures stability of cluster representatives within the competency space. Analytical representation of cluster center adjustment convergence is defined through Equation (58).

$$\Delta C = \max_k \| C_k^* - C_k^{(prev)} \| \qquad (58)$$

In Equation (58), the symbol ΔC denotes the maximum variation between optimized cluster centers and cluster centers obtained during the previous iteration. The vector $C_k^{(prev)}$ represents the cluster center position before optimization refinement. Minimal values of ΔC indicate stabilization of cluster representatives within the multidimensional competency environment.

3.10. Iterative Convergence Verification of Clustering Process

Iterative convergence verification ensures stabilization of clustering outcomes generated through density-adaptive cluster center optimization. Recruitment analytics environments frequently involve evaluation procedures that repeatedly refine candidate categorization until stable competency groupings emerge. Analytical verification of convergence measures variation between successive cluster center updates obtained from optimization cycles. Stabilization of cluster structures indicates that candidate competency patterns have reached consistent representation within the multidimensional feature environment. Convergence measurement is formulated through Equation (59), which evaluates the magnitude of variation between current and previous cluster centers.

$$\Omega_c = \frac{1}{Kp} \sum_{k=1}^{K} \sum_{j=1}^{p} | c_{kj}^* - c_{kj}^{(prev)} | \qquad (59)$$

In Equation (59), the symbol Ω_c denotes the convergence indicator describing average variation across cluster centers. The quantity c_{kj}^* represents the optimized value of the j^{th} attribute associated with the k^{th} cluster center. The term $c_{kj}^{(prev)}$ corresponds to the cluster center value obtained during the previous iteration. The parameter K indicates the number of clusters, while $p$denotes the number of attributes.

Convergence verification also requires examination of fuzzy membership stability across candidate records. Stable membership distributions indicate that candidate association with competency clusters remains consistent across successive clustering iterations. Analytical representation of membership stabilization is expressed through Equation (60).

$$\Omega_m = \frac{1}{mK} \sum_{i=1}^{m} \sum_{k=1}^{K} | \tilde{\mu}_{ik}^{(t)} - \tilde{\mu}_{ik}^{(t-1)} | \qquad (60)$$

In Equation (60), the symbol Ω_m represents the membership stabilization indicator. The value $\tilde{\mu}_{ik}^{(t)}$ denotes the reinforced membership value associated with candidate record i and cluster k during the current iteration, while $\tilde{\mu}_{ik}^{(t-1)}$ represents the membership value obtained during the previous iteration. The variable m corresponds to the number of candidate profiles contained within the dataset.

Completion of clustering iterations depends on a convergence decision criterion that evaluates stabilization indicators obtained from cluster center and membership analysis. Analytical formulation of the convergence condition is expressed through Equation (61), which determines termination of iterative clustering cycles.

$$\Lambda = \max(\Omega_c, \Omega_m) \quad (61)$$

In Equation (61), the symbol Λ represents the final convergence indicator derived from cluster center variation and membership stabilization measures. Smaller values of Λ indicate stable clustering behavior and confirm that candidate competency structures have achieved analytical equilibrium within the clustering environment.

Following convergence verification, candidate records receive final cluster assignments according to maximum reinforced membership values. Recruitment analytics interpretation resembles human-resource evaluation where candidate profiles become associated with competency groups reflecting strongest skill alignment. Analytical formulation of candidate assignment is represented through Equation (62).

$$\kappa_i = \arg\max_k (\tilde{\mu}_{ik}) \quad (62)$$

In Equation (62), the symbol κ_i represents the final cluster index assigned to the i^{th} candidate profile. The quantity $\tilde{\mu}_{ik}$ corresponds to the reinforced fuzzy membership value representing association between candidate record $i$and cluster center k.

Cluster profiling generates descriptive competency representations for each cluster based on density-weighted feature distributions of candidate records assigned to that cluster. Recruitment analytics environments interpret such profiles as representative skill patterns observed among groups of applicants. Analytical construction of competency cluster profiles is expressed through Equation (63).

$$P_{kj} = \frac{1}{N_k} \sum_{i \in C_k} \Gamma_{ij} \quad (63)$$

In Equation (63), the symbol P_{kj} represents the competency profile value for the j^{th} attribute associated with cluster k. The quantity Γ_{ij} denotes the density-weighted attribute value for the i^{th} candidate record. The set C_k represents candidate records assigned to cluster k, while N_k denotes the number of candidates within that cluster.

Final evaluation of clustering performance examines compactness and separation characteristics across competency clusters. Recruitment analytics frameworks rely on such measurements to determine the effectiveness of candidate grouping strategies. Analytical representation of cluster quality measurement is expressed through Equation (64).

$$Q_c = \frac{\sum_{k=1}^{K} \sum_{i \in C_k} \parallel \Gamma_i - P_k \parallel}{\sum_{k=1}^{K} N_k} \quad (64)$$

In Equation (64), the symbol Q_c represents the cluster quality indicator measuring compactness of candidate profiles around competency cluster centers. The vector Γ_i denotes the density-weighted feature vector representing candidate i, while P_k represents the competency profile vector associated with cluster k. The parameter N_k indicates the number of candidates assigned to the cluster.

4. DATASET SUMMARY

The Recruitment Data dataset available through Kaggle serves as a structured resource for analytical investigation of recruitment decision patterns. The dataset supports research related to candidate evaluation, hiring prediction, and workforce analytics within data-driven recruitment environments. Structured tabular records capture multiple attributes representing applicant characteristics, assessment outcomes, and organizational hiring responses. Availability of clearly organized variables enables systematic experimentation with machine learning and deep learning models focused on recruitment analytics.

The dataset contains thousands of candidate records collected from simulated or organizational hiring scenarios. Each record represents a unique candidate profile containing demographic information, qualification indicators, evaluation scores, and recruitment outcomes. Structured attributes include Candidate ID, gender category, academic qualification level, professional experience measured in years, test score obtained in screening examinations, interview performance score, communication skill rating, and domain knowledge assessment. Additional attributes describe employment status and final hiring decision assigned through recruitment evaluation. Such multidimensional attributes allow comprehensive examination of relationships between candidate characteristics and recruitment outcomes.

Quantitative attributes within the dataset enable statistical modeling and predictive analysis. Experience level, examination score, and interview performance values provide measurable indicators of professional capability. Communication skill rating and domain knowledge evaluation capture qualitative competence translated into numerical representation. Academic qualification level provides insight into educational background diversity across candidate profiles. Gender attribute enables demographic analysis related to recruitment fairness and representation across hiring outcomes. Hiring decision attribute functions as a target variable for classification tasks within predictive recruitment models.

Structured organization of variables supports multiple analytical objectives including candidate clustering, feature importance evaluation, and hiring probability prediction. Machine learning algorithms benefit from well-defined feature representation across candidate records. Deep learning architectures benefit from multidimensional attribute relationships that reveal complex recruitment patterns. Recruitment analytics models trained using this dataset support investigation of candidate suitability prediction, ranking frameworks for applicant prioritization, and fairness assessment across demographic attributes. Research utilization of the Recruitment Data dataset contributes toward development of intelligent decision support systems within human resource management. Analytical exploration of candidate attributes strengthens understanding of recruitment dynamics across qualification level, professional experience, and evaluation performance. Data-driven recruitment modeling derived from this dataset promotes transparent, consistent, and efficient hiring analysis across modern organizational environments.

5. RESULTS AND DISCUSSION

Results and discussions present systematic interpretation of experimental outcomes generated through comparative model evaluation. Analytical interpretation focuses on classification capability demonstrated by different computational frameworks using recruitment data containing 10,000 instances.

5.1. Classification Accuracy Analysis

Classification accuracy represents the proportion of correctly identified instances within the total evaluated observations, computed through the combination of true positive and true negative outcomes relative to the entire dataset. Higher classification accuracy reflects stronger predictive capability and improved model reliability in distinguishing correct decision boundaries. Table 1 holds the quantitative values obtained for the evaluated models, while Fig 1 visually represents comparative accuracy variation across the tested frameworks.

Table 1. Classification Accuracy

Classification Algorithms	Classification Accuracy
TransparentFit-AI	52.91
GraphHire-Net	62.22
FADW-KM	96.47

TransparentFit-AI records classification accuracy of 52.91%, indicating moderate predictive capability within the evaluated recruitment dataset. Observed performance level reflects balanced but limited decision discrimination, supported by balanced accuracy of 52.962 and moderate precision and recall values. GraphHire-Net improves predictive reliability with classification accuracy of 62.22%. Increased value demonstrates stronger capability in identifying relevant candidate outcomes, supported by improved sensitivity of 63.157 and specificity of 61.330. Reduction in false positive rate and false negative rate further strengthens decision consistency within the GraphHire-Net model, producing higher Matthews Correlation Coefficient and improved detection rate compared with TransparentFit-AI.

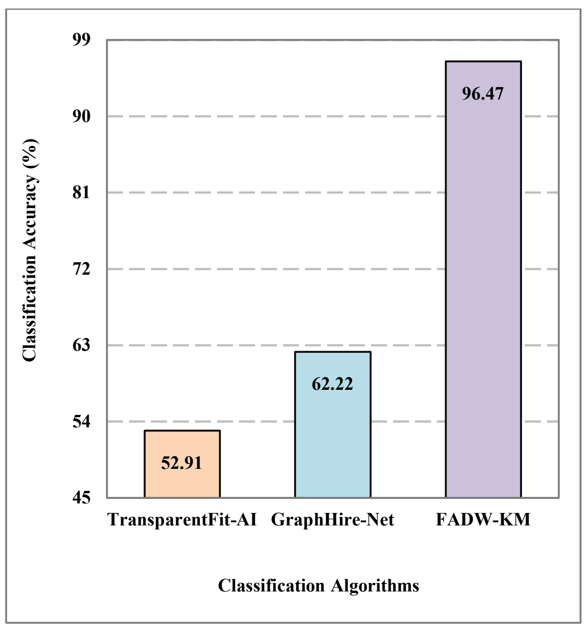

Fig 1. Classification Accuracy

FADW-KM exhibits a substantial performance advantage, achieving classification accuracy of 96.47% as shown in Table 1 and Fig 1. Observed value indicates significantly improved predictive precision within the evaluated recruitment dataset. Elevated accuracy demonstrates strong capability in correctly identifying both positive and negative classification outcomes. Enhanced detection reliability contributes to stronger decision consistency and improved classification confidence across the entire dataset. Comparative analysis clearly demonstrates progressive improvement from TransparentFit-AI to GraphHire-Net and finally to FADW-KM, indicating increasing model effectiveness in predictive recruitment analytics under identical dataset conditions.

5.2. Matthews Correlation Coefficient Analysis

Matthews Correlation Coefficient evaluates the relationship between predicted outcomes and actual class labels by incorporating true positives, true negatives, false positives, and false negatives within a single correlation measure. The coefficient value ranges from −1 to +1, where values close to +1 represent strong positive predictive correlation, values near zero indicate weak predictive association, and negative values represent inverse prediction behavior. Table 2 holds the computed Matthews Correlation Coefficient values for the evaluated algorithms, while Figure 2 presents graphical comparison of correlation strength among the classification frameworks.

Table 2. Matthews Correlation Coefficient

Classification Algorithms	Matthews Correlation Coefficient (%)
TransparentFit-AI	5.93
GraphHire-Net	24.48
FADW-KM	92.94

TransparentFit-AI records Matthews Correlation Coefficient value of 5.93%. Observed value remains very close to zero, indicating extremely weak correlation between predicted recruitment outcomes and actual dataset labels. Such a low coefficient suggests that predictive outputs generated by TransparentFit-AI demonstrate limited consistency in identifying correct candidate classification patterns. Weak correlation strength indicates reduced reliability in capturing structural relationships present within the recruitment dataset. GraphHire-Net demonstrates noticeable improvement by achieving Matthews Correlation Coefficient value of 24.48%. Increased value indicates moderate strengthening of predictive association compared with TransparentFit-AI. Improved coefficient suggests that GraphHire-Net captures certain structural relationships within candidate data more effectively, resulting in better alignment between predicted labels and true dataset classifications.

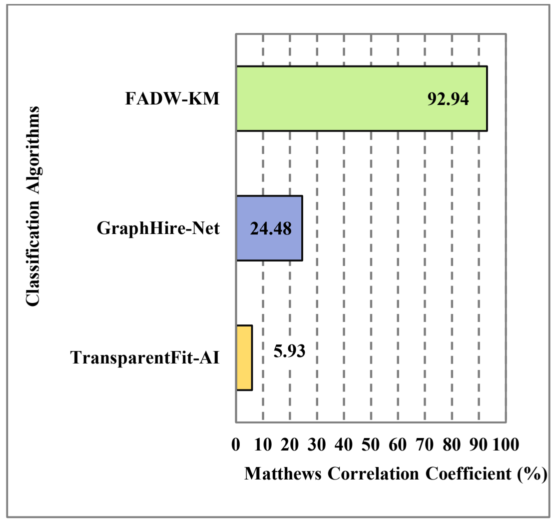

Fig 2. Matthews Correlation Coefficient

FADW-KM achieves Matthews Correlation Coefficient value of 92.94%, representing a substantial improvement in predictive correlation strength. Observed coefficient approaches the upper bound of the correlation scale, indicating extremely strong agreement between predicted classification results and actual recruitment dataset labels. High coefficient value reflects strong capability in maintaining balanced prediction performance across both positive and negative classes while minimizing classification inconsistencies. Comparative analysis illustrated in Table 2 and Figure 2 clearly indicates progressive improvement in correlation strength across the evaluated models, with FADW-KM demonstrating the highest predictive reliability and strongest classification consistency within the experimental recruitment dataset.

5.3. F1 Score Analysis

F1 Score represents a widely accepted evaluation measure that balances precision and recall in classification studies. The metric is calculated as the harmonic mean of precision and recall, ensuring balanced assessment between the ability to correctly identify positive instances and the capability to reduce false detections. A higher F1 Score indicates improved equilibrium between sensitivity and predictive precision, reflecting stable classification behavior within complex datasets. Table 3 holds the obtained F1 Score values for the evaluated algorithms, while Figure 3 illustrates the comparative variation in performance among the models.

TransparentFit-AI records an F1 Score of 49.32. Observed value indicates moderate balance between precision and recall within the evaluated recruitment dataset. The score suggests that the model demonstrates nearly equivalent capability in detecting relevant candidate outcomes and controlling incorrect positive predictions. Despite this balance, the overall magnitude remains limited, indicating restricted effectiveness in managing both classification sensitivity and precision simultaneously. Such a result reflects moderate predictive consistency within recruitment decision analysis.

Table 3. F1 Score

Classification Algorithms	F1 Score
TransparentFit-AI	49.32
GraphHire-Net	49.33
FADW-KM	49.90

GraphHire-Net achieves an F1 Score of 49.33. The value shows only marginal improvement over TransparentFit-AI, indicating that the harmonic balance between precision and recall remains largely similar. The slight increase reflects marginal refinement in prediction equilibrium, suggesting that GraphHire-Net produces slightly improved coordination between true positive identification and false detection control. Nevertheless, the difference remains extremely small, demonstrating that predictive stability across the two models remains almost identical within the evaluated dataset.

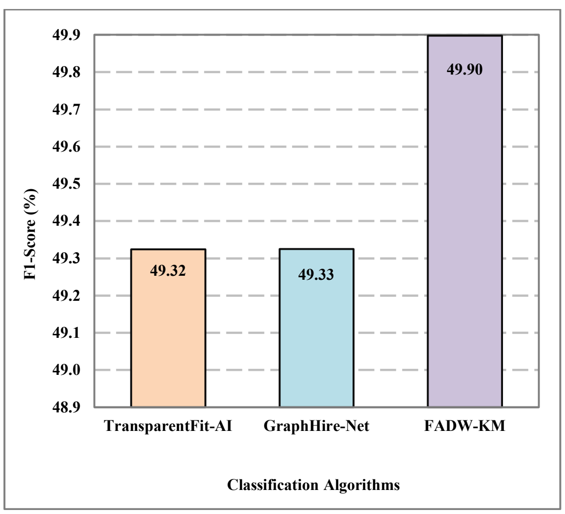

Fig 3. F1 Score

FADW-KM records the highest F1 Score value of 49.90. The increased score indicates comparatively stronger balance between precision and recall, reflecting improved coordination between sensitivity and classification precision during candidate prediction tasks. Although numerical improvement appears modest, the observed increase represents enhanced capability in managing both positive detection and false prediction control within the recruitment dataset. Comparative observation presented in Table 3 and Figure 3 demonstrates gradual improvement in F1 Score performance across the evaluated algorithms, indicating progressive enhancement in balanced classification behavior under identical experimental conditions.

5.4. G-Mean Analysis

G-Mean represents an important performance indicator used to measure the balance between sensitivity and specificity in binary classification studies. The metric is computed as the geometric mean of recall and specificity, ensuring equal consideration of correct positive detection and accurate negative classification. Higher G-Mean values indicate stronger capability in maintaining balanced prediction performance across both classes, reducing bias toward any single class category. Table 4 contains the calculated G-Mean values for the evaluated algorithms, while Figure 4 illustrates comparative variation in performance among the classification frameworks.

Table 4. G-Mean

Classification Algorithms	G-Mean
TransparentFit-AI	52.92
GraphHire-Net	62.24
FADW-KM	96.47

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

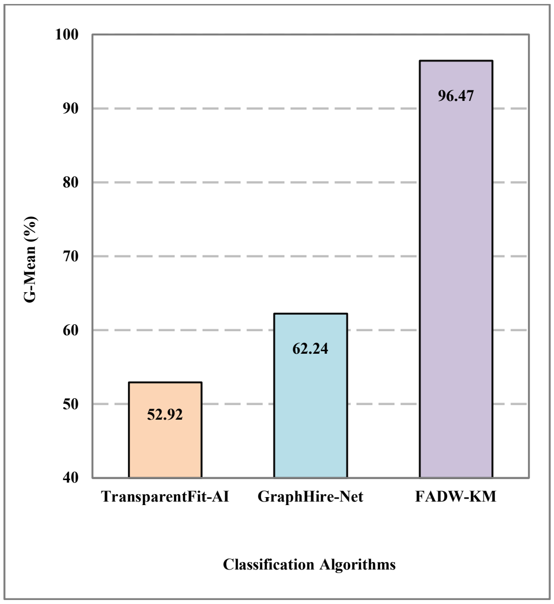

Fig 4. G-Mean

TransparentFit-AI records a G-Mean value of 52.92. The observed value indicates moderate balance between sensitivity and specificity within the evaluated recruitment dataset. Performance suggests that the model demonstrates limited capability in maintaining equal prediction reliability across positive and negative classification categories. Moderate geometric mean value indicates partial effectiveness in controlling classification imbalance, resulting in restricted predictive stability during recruitment outcome identification.

GraphHire-Net achieves a G-Mean value of 62.24. Increased value demonstrates noticeable improvement in classification balance compared with TransparentFit-AI. Enhanced geometric mean indicates improved capability in correctly identifying both relevant candidate instances and non-relevant outcomes within the dataset. The improvement suggests stronger stability in classification decisions, reflecting more reliable prediction behavior under recruitment analytics conditions.

FADW-KM records a substantially higher G-Mean value of 96.47. Observed value indicates extremely strong balance between sensitivity and specificity, demonstrating highly stable classification capability across both positive and negative recruitment categories. Elevated geometric mean reflects strong predictive consistency and effective control over classification imbalance within the dataset. Comparative evaluation presented in Table 4 and Figure 4 clearly demonstrates progressive improvement in classification balance across the evaluated models, with FADW-KM achieving the highest predictive stability and most reliable recruitment classification performance.

5.5. Critical Success Index Analysis

Critical Success Index represents a performance metric used to measure the proportion of correctly predicted positive outcomes relative to the total number of predicted and actual positive instances. The index evaluates the intersection between true positive predictions and the combined influence of false positives and false negatives. Higher Critical Success Index values indicate stronger capability in identifying correct positive classifications while minimizing prediction errors. Table 5 contains the calculated Critical Success Index values for the evaluated algorithms, and Fig 5 illustrates the comparative performance variation among the classification frameworks.

Table 5. Critical Success Index

Classification Algorithms	Critical Success Index (%)
TransparentFit-AI	36.25
GraphHire-Net	44.89
FADW-KM	93.23

TransparentFit-AI records a Critical Success Index value of 36.25%. Observed value indicates limited predictive success in identifying correct positive recruitment outcomes within the dataset. The moderate index reflects the presence of notable prediction inconsistencies influenced by both false positive and false negative classifications. Such performance suggests that the framework demonstrates restricted capability in capturing accurate candidate identification patterns, resulting in reduced effectiveness in recruitment outcome prediction.

GraphHire-Net achieves a Critical Success Index value of 44.89%. The improved value demonstrates stronger capability in identifying correct positive classifications compared with TransparentFit-AI. Higher index value indicates improved alignment between predicted positive instances and actual dataset outcomes. Reduction in incorrect classifications contributes to increased predictive reliability, suggesting that the GraphHire-Net framework captures relevant recruitment data relationships with greater effectiveness.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

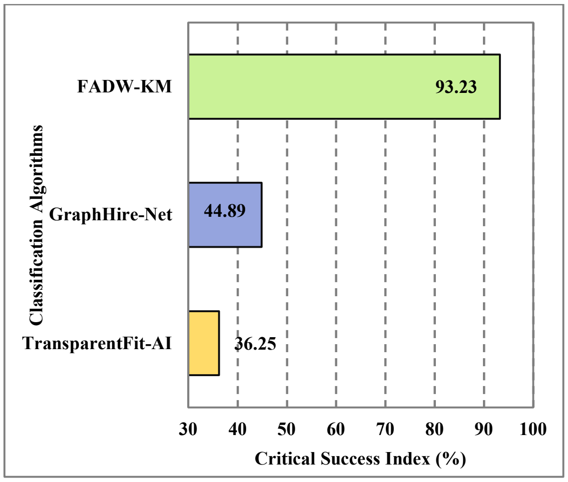

Fig 5. Critical Success Index

FADW-KM records a significantly higher Critical Success Index value of 93.23%. The elevated index reflects strong capability in accurately identifying positive recruitment outcomes while minimizing prediction errors associated with false detections. Such a high value indicates highly reliable classification behavior, demonstrating superior capability in recognizing correct candidate patterns within the recruitment dataset. Improved predictive alignment contributes to stronger classification consistency and decision reliability.

Comparative observation presented in Table 5 and Fig 5 highlights a clear progression in predictive success across the evaluated models. TransparentFit-AI demonstrates moderate performance, GraphHire-Net shows noticeable improvement, and FADW-KM achieves substantially higher predictive effectiveness, indicating superior capability in accurate recruitment classification analysis within the evaluated dataset.

Comparative evaluation across multiple performance metrics highlights consistent improvement in classification stability, correlation strength, and predictive balance. The framework demonstrates reliable classification behaviour across all evaluated metrics, indicating strong analytical capability in recruitment data analysis.

Recent studies in recruitment analytics emphasize recommendation systems and explainable decision models. Observed findings align with these approaches in improving predictive consistency, while extending analytical capability toward competency clustering and structured candidate grouping. The study considers a single recruitment dataset, which may limit generalization across diverse recruitment environments. Attribute representation relies on structured features without deep semantic interpretation. Computational complexity may increase with large-scale high-dimensional data.

6. CONCLUSION AND FUTURE WORKS

The study presents a structured recruitment analytics framework designed to improve candidate classification and competency-based grouping within large-scale recruitment datasets. The research objective focused on developing an analytical model capable of capturing heterogeneous candidate characteristics through density-aware clustering, adaptive attribute weighting, and flexible membership representation. The methodological design enabled systematic transformation of recruitment data into competency-oriented clusters, supporting reliable interpretation of candidate structures and relationships.

The proposed framework integrates density-sensitive analysis with weighted similarity evaluation to enhance the identification of meaningful candidate groupings. Fuzzy membership modelling supports representation of overlapping competency profiles, reflecting real-world recruitment scenarios where candidate attributes do not follow rigid boundaries. Iterative cluster refinement ensures stable convergence and consistent analytical outcomes across the dataset. Such integration strengthens classification behaviour and improves the reliability of recruitment decision analysis by preserving both local data distribution and global structural relationships.

Experimental evaluation demonstrates improved classification consistency, balanced predictive behaviour, and strong correlation between predicted and actual candidate outcomes across multiple performance metrics. Comparative analysis against existing recruitment models confirms enhanced capability in capturing competency patterns and maintaining stable classification performance. Observed improvements indicate effective alignment between the proposed methodology and the research objectives, validating the suitability of the framework for structured recruitment analytics.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The developed framework contributes to recruitment data analysis by providing a systematic approach for competency clustering and candidate evaluation. Analytical design supports improved understanding of candidate distributions and enables structured interpretation of recruitment data. Such capability is valuable for data-driven recruitment environments requiring consistent and reliable analytical support for candidate selection processes. Future enhancements may consider integration of semantic feature extraction and large-scale deployment strategies to further strengthen analytical capability and applicability across diverse recruitment domains.

REFERENCES:

[1] F. Mentec, Z. Miklos, S. Hervieu, and T. Roger, “AI4HR Recruiter: a job recommender system for internal recruitment in consulting companies,” *Procedia Comput. Sci.*, vol. 225, pp. 1231–1240, 2023, doi: https://doi.org/10.1016/j.procs.2023.10.111.

[2] A. Oliveira *et al.*, “Revolutionizing Recruitment with AI: Empirical Insights into Benefits and Challenges,” *Procedia Comput. Sci.*, vol. 263, pp. 696–703, 2025, doi: https://doi.org/10.1016/j.procs.2025.07.083.

[3] S. Pokharna, P. Sharma, T. Taneja, S. Verma, and H. Ojha, “Secure & Reliable Fake Profile Detection on Recruitment Platforms using Machine Learning and Blockchain,” *Procedia Comput. Sci.*, vol. 259, pp. 1228–1238, 2025, doi: https://doi.org/10.1016/j.procs.2025.04.078.

[4] Z. Wang, W. Wei, C. Xu, J. Xu, and X.-L. Mao, “Person-job fit estimation from candidate profile and related recruitment history with co-attention neural networks,” *Neurocomputing*, vol. 501, pp. 14–24, 2022, doi: https://doi.org/10.1016/j.neucom.2022.06.012.

[5] S. Lee and T. Jang, “Advancing water level prediction using clustering-based machine learning techniques in data-scarce regions,” *Environmental Modelling & Software*, vol. 198, p. 106899, 2026, doi: https://doi.org/10.1016/j.envsoft.2026.106899.

[6] H. Huang, S. Zuo, Y. Lu, and S. Wang, “Brand Consumption Potential Rating Model for Business Cluster Based on FAHP,” *Procedia Comput. Sci.*, vol. 261, pp. 1026–1035, 2025, doi: https://doi.org/10.1016/j.procs.2025.04.681.

[7] A. Ghimire, M. Alkurdi, S. Mohanty, and F. Amsaad, “AI-enabled image processing approach for efficient clustering and identification of hardware Trojans,” *Integration*, vol. 107, p. 102628, 2026, doi: https://doi.org/10.1016/j.vlsi.2025.102628.

[8] M. Cavus, M. N. Benli, U. Altuntas, M. Sari, H. Ayan, and Y. F. Ugurluoglu, “Transparent and bias-resilient AI framework for recidivism prediction using deep learning and clustering techniques in criminal justice,” *Appl. Soft Comput.*, vol. 176, p. 113160, 2025, doi: https://doi.org/10.1016/j.asoc.2025.113160.

[9] A. A. Rachman and M. I. Sofyan, “An Analysis of Indonesia’s Online Job Recruitment Platform Comparing the Benefits and Costs of AI Meta-Ads with and without AI,” *Procedia Comput. Sci.*, vol. 269, pp. 21–28, 2025, doi: https://doi.org/10.1016/j.procs.2025.08.255.

[10] C. P. Contiguglia, G. Quaranta, C. Demartino, and B. F. Spencer, “AI-driven conceptual optimization of building façade layouts using a fuzzy-logic-based morphological index,” *Autom. Constr.*, vol. 182, p. 106750, 2026, doi: https://doi.org/10.1016/j.autcon.2025.106750.

[11] A. Jalali, H. Shariatmadar, and S. Golnargesi, “Probabilistic active control of a seismically excited building using probabilistic fuzzy logic controller,” *Soil Dynamics and Earthquake Engineering*, vol. 203, p. 110131, 2026, doi: https://doi.org/10.1016/j.soildyn.2026.110131.

[12] K. Konar, S. Das, S. Das, and S. Misra, “Employee Attrition Prediction Using Bayesian Optimized Stacked Ensemble Learning and Explainable AI,” *SN Comput. Sci.*, vol. 6, no. 6, p. 672, 2025, doi: 10.1007/s42979-025-04204-w.

[13] D. Puthal, A. K. Mishra, S. P. Mohanty, A. Longo, and C. Y. Yeun, “Shadow AI: Cyber Security Implications, Opportunities and Challenges in the Unseen Frontier,” *SN Comput. Sci.*, vol. 6, no. 5, p. 405, 2025, doi: 10.1007/s42979-025-03962-x.

[14] L. Y. Cabrera, J. Wagner, S. Gerke, and D. Susser, “Tempered enthusiasm by interviewed experts for synthetic data and ELSI checklists for AI in medicine,” *AI and Ethics*, vol. 5, no. 3, pp. 3241–3254, 2025, doi: 10.1007/s43681-024-00652-x.

[15] I. Mainardi, “Change management: artificial intelligence (AI) at the service of public administrations,” *AI Soc.*, vol. 40, no. 5, pp.

3953–3981, 2025, doi: 10.1007/s00146-024-02136-2.

[16] H. Javed, S. El-Sappagh, and T. Abuhmed, “Robustness in deep learning models for medical diagnostics: security and adversarial challenges towards robust AI applications,” *Artif. Intell. Rev.*, vol. 58, no. 1, p. 12, 2024, doi: 10.1007/s10462-024-11005-9.

[17] G. Cartocci *et al.*, “Explainable artificial intelligence in air traffic control: effects of expertise on workload, acceptance, and usage intentions,” *Brain Inform.*, vol. 13, no. 1, p. 6, 2026, doi: 10.1186/s40708-025-00287-6.

[18] V. Terziyan *et al.*, “Towards ethical evolution: responsible autonomy of artificial intelligence across generations,” *AI and Ethics*, vol. 5, no. 5, pp. 5165–5190, 2025, doi: 10.1007/s43681-025-00759-9.

[19] A. C. Tran and N. C. Tran, “A recruitment support system based on Large Language Models and Retrieval-Augmented Generation,” *Procedia Comput. Sci.*, vol. 269, pp. 259–268, 2025, doi: https://doi.org/10.1016/j.procs.2025.08.278.

[20] L. Gonzalez-Garcia, M.-A. Sicilia, and E. García-Barriocanal, “Classification of Job Offers into Job Positions Using O∗NET and BERT Language Models,” *Computers, Materials and Continua*, vol. 86, no. 2, pp. 1–15, 2025, doi: https://doi.org/10.32604/cmc.2025.070813.

[21] F. Wu, A. Ye, Z. Cai, and J. Chen, “Towards bias-free recruitment: Adversarial contrastive tuning for LLM-based resume screening,” *Inf. Softw. Technol.*, vol. 193, p. 108040, 2026, doi: https://doi.org/10.1016/j.infsof.2026.108040.

[22] E. Albaroudi, T. Mansouri, M. Hatamleh, and A. Alameer, “HitHire: The future of ethical, fair, and sustainable AI recruitment – A governance framework,” *Array*, vol. 29, p. 100592, 2026, doi: https://doi.org/10.1016/j.array.2025.100592.

[23] M.-H. Ajjam and H. S. Al-Raweshidy, “AI-driven semantic similarity-based job matching framework for recruitment systems,” *Inf. Sci. (N. Y).*, vol. 724, p. 122728, 2026, doi: https://doi.org/10.1016/j.ins.2025.122728.

[24] A. Chen, F. Han, X. Zhang, and Y. Lu, “Cracking the AI recruitment code : Striving for transparency in finding the right person – job fit,” *Information & Management*, vol. 62, no. 5, p. 104156, 2025, doi: 10.1016/j.im.2025.104156.

[25] X. Shi, C. Wang, and Q. Wei, “ProMatch : A novel dynamic process-unpacking approach for two-way proactive recruitment,” *Decis. Support Syst.*, vol. 200, no. November 2025, p. 114564, 2026, doi: 10.1016/j.dss.2025.114564.

[26] M. P. Swapna, D. Rajeev, J. Ramkumar, and S. Chandran, “Unveiling Cybercrime Patterns in Kerala: A Machine Learning Approach,” in Lecture Notes in Networks and Systems, vol. 1445, 2026, pp. 111–122.

[27] J. Ramkumar and S. Vetrivel, “Deep Learning Insights into Defending Against Adversarial Attacks in IoT Systems,” in Strategic Approaches to Intrusion Detection in Cloud IoT Ecosystem, 2026, pp. 227–250.

[28] R. Suganthi, J. Ramkumar, K. K. R. Chinthala, M. Deenathayalan, and K. R. Dhanya, “Real-Time Applications of 6G in Augmented Reality and Virtual Reality Ecosystems,” in Landscaping 6G Unlocking the Power of Ultra Fast Communication, 2026, pp. 175–186.

[29] B. Suchitra, J. Ramkumar, and V. Valarmathi, “Rejuvenated Particle Swarm Optimization-Based Classifier for Big Sentiment Data Classification,” in Communications in Computer and Information Science, vol. 2819, 2026, pp. 41–54.

[30] V. Valarmathi, J. Ramkumar, and B. Suchitra, “Classification of Cross-Domain Sentiments in Big Data Using Renewed Genetic Algorithm,” in Communications in Computer and Information Science, vol. 2819, 2026, pp. 25–40.

[31] J. Joy, S. Devaraju, and J. Ramkumar, “Utilizing Fuzzy-Identity-Based Encryption with Proxy-Re-Encryption for Data Sharing,” Journal of Theoretical and Applied Information Technology, vol. 103, no. 18, pp. 7469–7479, 2025.

[32] J. Ramkumar and V. Valarmathi, “Harnessing AI-Driven Models for Sustainable Development in Business Management,” in World Sustainability Series, 2025, pp. 217–238.

[33] R. Jaganathan, S. Rajagopal, and K. Rajendran, “Cultural Intelligence in the AI Era Enhancing Transitional Higher Education,” in Bridging Global Divides for Transnational Higher Education in the AI Era, 2024, pp. 273–292.

[34] J. Ramkumar, B. Narasimhan, S. P. Priyadharshini, and R. Karthikeyan, “Optimizing Intrusion Detection with Bayesian Optimization in Reactive Intrusion

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Detection Systems for Enhanced Security Performance," in Artificial Intelligence Based Smart and Secured Applications, S. Rajagopal, P. Sajja, R. Thanki, and A. Kumar, Eds., Communications in Computer and Information Science, vol. 2822, Cham, Switzerland: Springer, 2026.

[35] R. Karthikeyan, J. Ramkumar, S. P. Priyadharshini, and B. Narasimhan, "Optimized Self-Adaptive Intrusion Detection System for Enhanced Network Security," in Artificial Intelligence Based Smart and Secured Applications, S. Rajagopal, P. Sajja, R. Thanki, and A. Kumar, Eds., Communications in Computer and Information Science, vol. 2822, Cham, Switzerland: Springer, 2026.