ISSN: 2788–7669 *Journal of Machine and Computing 5(3)(2025)*

Artificial Intelligence Based Recruitment Prediction and Sentiment Analysis for Enhanced HR Efficiency

[1]Mano Ashish Tripathi, [2]Dhanalakshmi Komatiguntala, [3]Sree Lakshmi Moorthygari, [4]Sundari Dadhabai, [5]Amit Mishra and [6]Ravi Kumar Bommisetti
[1]School of Management Studies, Motilal Nehru National Institute of Technology Allahabad, Prayagraj, Uttar Pradesh, India.
[2]School of Business and Management, Christ University, Yeswanthpur Campus, Bangalore, India.
[3]Department of Business Management, Mahatma Gandhi University, Nalgonda, Telangana, India.
[4]KL Business School, Koneru Lakshmaiah Educational Foundation, Vaddeswaram, Andhra Pradesh, India.
[5]Department of Computer Science and Applications, Dr. Vishwanath Karad, MITWPU, Pune, Maharashtra, India.
[6]Independent Researcher, Andhra Pradesh, India.
[1]manoashish@mnnit.ac.in, [2]dhanalakshmi.k@christuniversity.in, [3]sreelakshmi.mgu@gmail.com, [4]sundaridadhabai@kluniversity.in, [5]i.amitmishra@gmail.com, [6]ravi9949418650@yahoo.com

Correspondence should be addressed to Ravi Kumar Bommisetti : ravi9949418650@yahoo.com

ArticleInfo
Journal of Machine and Computing (https://anapub.co.ke/journals/jmc/jmc.html)
Doi : https://doi.org/10.53759/7669/jmc202505145.
Received 16 January 2025; Revised from 07 May 2025; Accepted 17 June 2025.
Available online 05 July 2025.

Abstract – In the present era of data-driven organizational environment, the practice of Human Resource Management (HRM) has become increasingly reliant on intelligent Decision-Support Systems (DSS). This study develops a multifaceted two-pipeline model of Predictive Modelling (PM) and Sentiment Analysis (SA) to enhance workforce analytics capabilities. A publicly available HRM analytic dataset is used to train supervised classification models, including Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Long Short-Term Memory (LSTM), as well as an ensemble model that integrates these classifiers. These approaches use structured data to predict employee attrition based on features such as age, job role, experience, and job satisfaction. The unstructured textual data sources, including resumes and employee reviews, are handled using *state-of-the-art* Natural Language Processing (NLP) such as tokenization, Term Frequency-Inverse Document Frequency (TF-IDF), and Bidirectional Encoder Representations as Transformers (BERT)-based embeddings. The new Mathematically Modified Robustly Optimized BERT Pretraining (MM-RoBERTa) is proposed for extracting the PM and SA. All the models are evaluated using k-fold Cross-Validation (CV) and standard evaluation measures, namely Accuracy, F1-score, Area Under the Receiver Operating Characteristic Curve (AUC), and Mean Absolute Error (MAE). The ensemble model achieves a predictive accuracy of 91.3%, and MM-RoBERTa outperforms existing SA with an accuracy of 93.1 %. The combination of predictive and affective insights is of practical use in fine-tuning talent retention, empowering HRM professionals to make informed decisions based on objective performance indicators and subjective emotional states.

Keywords – Natural Language Processing, Predictive Modelling, Sentiment Analysis, Employee Attrition, Ensemble Learning, Workforce Analytics, Classification Performance.

I. INTRODUCTION

In the current competitive and digitally disrupted business environment, organizations face increasing pressure to attract, engage, and retain top talent efficiently [1]. The traditional recruitment practices—characterized by manual resume screening, subjective assessments [2], and Life Time Consumer (LTC) processes—often result in suboptimal hiring decisions and high operational costs. To address these challenges, Artificial Intelligence (AI) has been increasingly adopted in Human Resource Management (HRM). AI-based Predictive Modelling (PM) and Sentiment Analysis (SA) are transforming talent acquisition by enabling data-driven decisions, automating repetitive tasks, and enhancing candidate assessment accuracy [3].

Talent acquisition utilizing Artificial Intelligence (AI)-based PM employs Machine Learning (ML) algorithms to analyze historical hiring data, workforce performance data, and candidate profiles, thereby predicting the suitability and future performance of potential employees [4]. PM combines and correlates data across structured and unstructured data to generate ranked candidate lists based on predictive probabilities for success in the job, cultural fit, and probability of retention probability [5]. This data-centric method reduces dependence on intuition and mitigates cognitive bias. Commonly employed algorithms include Logistic Regression (LR), Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM), and Deep Learning (DL) such as Neural Networks (NN). Such models are trained using labeled datasets that contain attributes such as educational context, work experience, skill sets, interview performance scores, and post-hire productivity indices and are used to rank automatically and shortlist candidates [6].

SA—a subfield of Natural Language Processing (NLP)— is employed to extract subjective insights and emotional tone from textual data [7]. Regarding recruiting, SA may be used to evaluate candidate resumes, cover letters, email communications, social media presence, and interview transcripts. SA can aid in assessing candidate motivation, emotional intelligence, and fit based on their semantic features. This provides another qualitative dimension to the screening exercise, thereby helping recruiters identify soft skills and personality features that may not be evident in structured application data or other structured application information.

Integrating Predictive Modelling (PM) and SA into recruitment platforms proposes several tangible benefits [8]. It leads to the first one significantly enhancing operational efficiency by reducing time-to-hire through automated resume screening, intelligent candidate matching, and priority-based ranking mechanisms. Subsequently, it enhances the quality of hiring by improving talent identification and uncovering high-potential candidates who might be overlooked during conventional filtering [9]. Furthermore, it promotes objectivity and fairness, reducing the human bias factor and aligning evaluation expectations for all applicants. Continuous feedback mechanisms enable dynamic learning, aligning with organizational hiring trends and adapting to changing skill requirements.

AI can deliver real-time analytics dashboards to HRM managers, including recruitment pipeline performance, recruiter productivity, diversity metrics, and attrition risk analysis. The sentiment trends in candidate communications can reveal early indicators of disengagement or highlight opportunities for proactive engagement [10]. AI-driven chatbots and virtual assistants facilitate preliminary candidate interactions, providing answers to commonly requested information and even conducting initial interviews, offering a seamless and engaging experience to candidates.

Despite its transformative capabilities, AI in recruitment raises ethical and operational considerations that must be addressed. These problems represent algorithmic bias, data privacy, model explainability, and human oversight, which are urgent and require the implementation of robust, responsible AI governance frameworks [11]. To ensure fairness and legality, it is crucial to ensure that AI models are trained on unbiased, representative datasets and undergo regular audits.

AI-driven PM and SA represent a paradigm shift in talent acquisition, empowering HRM teams to make faster, data-driven, and unbiased hiring decisions [12]. As organizations increasingly adopt AI to address talent shortages and meet evolving workforce demands, the intersection between data science and human resource management is likely to become a strategic cornerstone, driving innovation, agility, and competitive advantage in the recruitment market.

II. RELATED WORKS

[12] Thoroughly review the uses of AI technologies in changing HRM, including Talent Acquisition (TA), development, and retention. The research classifies AI as ML, NLP, Robotic Process Automation (RPA), and Predictive Analytics (PA) as important facilitators of the optimization of HRM activities. Within the TA field, the authors emphasize that AI can be used to automatize the resume screening process, perform smart candidate searches, and enhance candidate engagement via chatbots and recommendation systems. Such innovations streamline hiring schedules and provide more accuracy in the process.

The paper also highlights the importance of AI in Talent Development (TD), where adaptive learning platforms and AI-powered performance monitoring systems tailor career advancement and training trajectories to the individual requirements of employees. Regarding Talent Retention (TR), Kadirov et al. explain how PA can be used to predict possible risks of attrition characterized by behavioral and past HR-related metrics, enabling organizations to deploy retention interventions before it is too late.

The Ethical Dimension (ED) is also actively covered. The authors are alerting to the potential harm of Algorithmic Bias (AB), the lack of explanation, and the risks of Data Privacy (DP), and are promoting the responsible use of AI. The analytical depth of the study is its strong point, as the visual illustrations and comparative models provide practical recommendations to HRM professionals and policymakers.

In parallel with it, [13] applies Computational Literature Review (CLR), namely BERTopic (Bidirectional Encoder Representations from Transformers for Topic Modelling), to the history of AI in Talent Management (TM). Their results support the Kadirov et al. perspective concerning the strategic impact of AI on HRM processes, namely recruitment, engagement, and Performance Management (PM). They observe that the potential of AI and HRM activities is becoming increasingly aligned, although there are methodological shortcomings.

[14] Support these findings by emphasizing the use of ML in candidate screening, employee assessment, and succession planning (SP). They draw attention to the application of classification models and PA in recognizing highly performing employees and increasing engagement. [15] confirm the effectiveness of DLs, Convolutional Neural Networks (CNNs), and Deep Neural Networks (DNNs) in recruitment and retention tasks and note the high values of accuracy and Area Under Curve (AUC), which additionally proves the theoretical propositions.

In general, the work of [16, 17] is an introductory survey that succeeds in setting the stage for AI-driven HRM as it is and as it should be. The research aligns well with recent publications that have demonstrated the practical usefulness and highlighted the ethical dilemmas of applying AI to the HRM workflow, particularly in talent acquisition and predictive workforce analytics.

Although the existing literature on AI in HRM offers several valuable insights, it has several weaknesses in the domains of talent acquisition (TA) and decision intelligence [18, 19]. The vast majority of research provides high-level overviews of AI, including ML, NLP, and PA, but does not offer models for implementing them end-to-end [20]. They also tend to view recruitment, development, and retention in silos, failing to create a coherent, data-driven talent lifecycle management pipeline.

The second significant gap is the minimal application of SA. Very little current research takes advantage of the SA by candidates or employees written in resumes, during interviews, or through feedback mechanisms, a lost opportunity to gauge cultural match, motivation, and emotional temperament. Additionally, there is limited exploration of dynamic modelling with real-time or multimodal data (Text, Demographics, Behavior), which reduces the PM.

Therefore, an opportunity to find an AI-based MP with NLP-based SA is evident. This would allow automated and bias-free SA recruitment DSS, enhance the candidate experience, and finetune retention plans. This is where PM and SA overlap: it is possible to considerably increase recruitment effectiveness, decrease turnover, and achieve strategic staffing planning, which is the gap in the existing intelligent HRM.

III. AI DRIVE MODEL FOR TA AND SA

The AI-TA model utilizes PM and SA to identify talent and enhance hiring accuracy and organizational fit. PM is a supervised ML using the RF, XGBoost, and Long Short-Term Memory (LSTM) to process structured HRM data: age, education, job position, experience, and satisfaction scores to predict employee turnover and appropriateness. Recursive Feature Elimination (RFE) and correlation-based feature elimination help eliminate irrelevant or redundant features, thus enhancing model robustness. The generalization is enhanced by the ensemble method that consists of heterogeneous learners and a logistic regression meta-learner. This pipeline is complemented by SA, which parses unstructured textual data, including resumes, feedback, and reviews, with the help of innovative NLP techniques such as tokenization, TF-IDF, and BERT embeddings. MM-RoBERTa is a proposed model pre-trained on a domain corpus and regularized with mathematical dropout to extract emotion and SA, which are used to evaluate behavioral features. Collectively, this mixed model can ensure data-driven, bias-minimized, and comprehensive recruiting by combining performance-based measures with effective ones. **Fig 1** provides the general model.

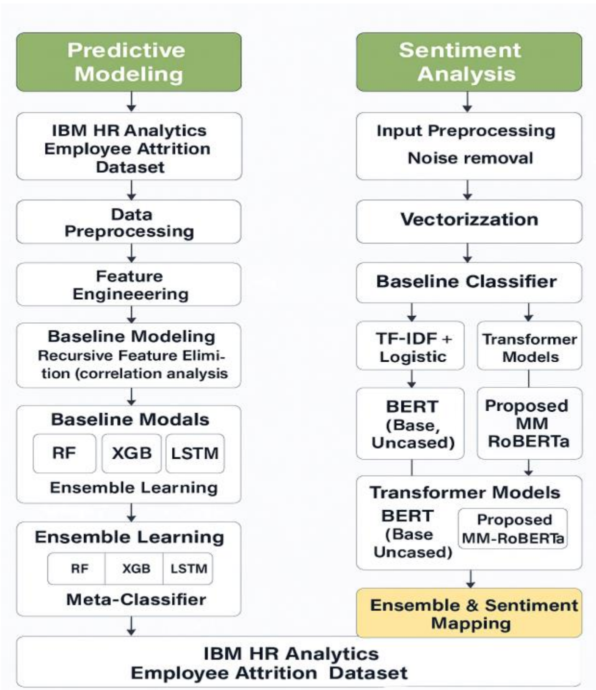

Fig 1. The Proposed Model: AI-Drive for TA and SA.

IV. PREDICTIVE MODELLING

This PM pipeline is proposed to model employee attrition and suitability decisions based on a structured dataset and an ensemble DL with a stacked ensemble architecture to maximize performance. The complete model is divided into consecutive steps: Data Preprocessing, Model Development, and Ensemble Learning.

Data Preprocessing

The IBM-HRM Analytics Employee Attrition dataset has been utilized, and the features include age, monthly income, job satisfaction, number of companies worked for years of working experience, and attrition status. Suppose matrix D∈Rn×m represents the dataset, n is the number of employees, and m is the number of features. Equation 1 is used to impute missing values as mean imputation.

$$x_{i,j} = \frac{1}{|N_j|} \sum_{x \epsilon N_j} x \tag{1}$$

Where,

- N_j→ The set of non-null values in column 'j'.

Categorical variables (*e.g.,* JobRole*,* Marital Status) are label-encoded or one-hot encoded. The process of label encoding is accomplished using Equation 2.

$$x_{i,j}^{(encoded)} = LabelMap(x_{i,j}) \tag{2}$$

Where,

- LabelMap→ A bijective function mapping type labels to integers.

Feature Engineering

The feature correlation matrix is generated using Equation 3.

$$C_{i,j} = \frac{Cov(x_i, x_j)}{\sigma_{x_i} \cdot \sigma_{x_j}} \tag{3}$$

- Highly correlated redundant features (|C| > 0.9) are removed to prevent multicollinearity. A stratified 80:20 split is employed to preserve class distribution for the binary target variable y∈{0,1}, where 1 indicates attrition.

Model Development

RF is an ensemble classifier that aggregates multiple DTs as T_1*,* T_2*,..., and* T_k trained on bootstrapped samples. The process is given using Equation 4.

$$\hat{y} = mode\{T_1(x), T_2(x), \dots \dots, T_k(x)\} \tag{4}$$

Each tree minimizes the Gini impurity using Equation 5.

$$G = \sum_{i=1}^{c} p_i(1 - p_i) \tag{5}$$

Where,

- p_i→ The probability of class '*I*' in a node.

XGBoost builds additive models using gradient-boosted DT. The formulation of XGBoost is given in Equation 6.

$$\hat{y_i} = \sum_{t=1}^{T} f_t(x_i), \quad f_t \in \mathrm{F} \tag{6}$$

Where,

- F→ The space of regression trees.

The objective function is given using Equation 7.

$$L = \sum_{i=1}^{n} l(y_i, \hat{y_i}) + \sum_{t=1}^{T} \Omega(\mathrm{f_t}) \tag{7}$$

The regularisation term is given in Equation 8.

$$\Omega(\mathrm{f}) = \gamma\mathrm{T} + \frac{1}{2}\lambda||w||^2 \tag{8}$$

LSTM is applied for modelling sequential features (*e.g.*, time in the company, job changes). The cell dynamics are indicated using Equations 9 to 14.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{9}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{10}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \widetilde{C_t} \tag{11}$$

$$\widetilde{C_t} = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{12}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{13}$$

$$h_t = o_t \odot \tanh(c_t) \tag{14}$$

Proposed Ensemble Learning

To leverage the strengths of all base learners, a stacked generalization is constructed with LR as a meta-learner.

Let $f_{RF(x)}$, $f_{XGB(x)}$, $f_{LSTM(x)}$ as base model predictions.

The meta-input vector is given by Equation 15.

$$Z = [f_{RF(x)}, f_{XGB(x)}, f_{LSTM(x)}] \tag{15}$$

The final prediction is given by Equation 16.

$$\hat{y} = \sigma(W^T Z + b) \tag{16}$$

Where,

- Σ→ The sigmoid function for binary classification that is given in Equation 17.

$$\sigma(z) = \frac{1}{1+e^{-z}} \tag{17}$$

The meta-learner is trained on out-of-fold predictions to prevent data leakage. The PM is explained using Algorithm 1.

Algorithm 1. PM Pipeline for Employee Attrition Prediction

Step 1: Data Preprocessing

- Load dataset D *# Structured HRM (e.g., IBM HR Analytics)*

For Each column in D:

If the column has missing values:

- Replace missing values with column mean

For Each definite feature in D:

If cardinality is low:

- Apply Label Encoding

Else:

- Apply One-Hot Encoding

For Each numerical feature in D:

- Standardize to ‘0’ mean and unit variance

Compute correlation matrix for all features

- Remove one of any pair of features with correlation>0.9
- Split D into *Train_Set()* and *Test_Set()* (80:20 stratified split on target)

Step 2: Model Development

- Initialize RF with *n_Estimators, Max_Depth().*
- Fit RF on *Train_Set()*
- Initialize XGB model with *Learning_Rate(), Max_Depth()*
- Fit XGB on *Train_Set* using early stopping
- Reshape sequential features as required
- Build LSTM
- Input→LSTM layer(s)→ Dense→ Output
- Compile with *Binary_Crossentropy()* loss and suitable optimizer
- Fit LSTM on *Train_Set()* with validation split

Step 3: Ensemble Learning via Stacking

- **Initialize** empty matrix Z for meta-input
- **For Each** *Base_Model* [RF, XGB, LSTM]:
- Perform k-fold cross-validation:
- **For Each** fold:
    - Train *Base_Model()* on training folds
    - Predict on validation fold
    - Store predictions in Z

Step 4: Initialize Logistic Regression model

- Fit on Z using true labels from validation sets

- Obtain RF, XGB, and LSTM predictions on Test_Set
- Stack predictions into vector z = [pred_RF, pred_XGB, pred_LSTM]
- y_Pred = *LogisticRegression.predict(z)*

Return y_Pred

Sentiment Analysis

The SA pipeline aims at retrieving emotion type and SA in unstructured HRM textual data, including resumes, performance feedback, and employee reviews. Such a pipeline is organized into three significant steps: Text Preprocessing, Model Development, and Emotion and Sentiment Mapping. They include traditional and transformer-based DL and an improvement in terms of MM-RoBERTa is presented.

- **MM-RoBERTa:** Mathematically Modified Robustly Optimized BERT enhances RoBERTa with the subsequent steps.

Instead of standard dropout, apply feature–importance–aware dropout using Equation 18.

$$x' = x \cdot m, m_i \sim \text{Bernoulli}\ (1 - \mathrm{p_i}) \tag{19}$$

Where,

$\mathrm{p_i}$→ Adaptive and inversely related to the feature importance score.

- **RoBERTa** is further pre-trained on a domain-specific HR corpus DHR using the Masked Language Modelling (MLM) objective specified in Equation 19.

$$\mathrm{L_{MLM}} = -\sum_{i \epsilon M} Log\ P(x_i|\hat{x}) \tag{20}$$

Where,

- M→ The set of masked positions and $'\hat{x}'$ is the masked input sequence.

To mitigate label imbalance, the use of weighted cross-entropy is specified in Equation 20.

$$L_{weighted} = -\sum_{i=1}^{N} w_{y_i}, \log P(y_i|x_i) \tag{21}$$

Where,

- $w_{y_{ij}} = \frac{1}{f_{y_i}}$, with f_{y_i} being the class frequency

The last hidden states of the baseline and transformer-based models are projected into SA space to make the models interpretable and capable of driving action. In the SA, the predicted output '*y*' is classified as positive when *y=2*, as neutral when *y=1,* and as negative when *y=0.* Parallel to this, multi-label emotion classification is carried out to identify emotions, including Joy, Anger, Fear, and Sadness. That is done with sigmoid activation functions, as the probability of each emotion label to appear is calculated separately.

The prediction is determined by the following for each emotion class: j.

$$P(y_i = 1|x) = \sigma(w_j^T \mathrm{x} + \mathrm{b_j}) \tag{22}$$

Where,

- x→ The input embedding vector
- $\mathrm{W_j}$, $\mathrm{b_j}$ → The class-specific parameters.

The model is optimized using binary cross-entropy loss across all emotion labels, allowing it to identify multiple co-occurring emotional states in textual input.

The SA combines classic (TF-IDF + LR) and deep transformer-based (BERT, MM-RoBERTa). MM-RoBERTa improves performance by using mathematically regularized dropout, domain adaptation with HRM-specific text, and class-balanced training. The multi-label emotion classification adds utility to subtle HRM analysis. The proposed pipeline can empower HRM professionals to extract structured information from unstructured inputs, enabling them to make better decisions during the hiring process, employee appraisals, and monitoring employee well-being in the workplace. The procedure of SA is given in Algorithm 2.

Algorithm 2. Sentiment and Emotion Mapping Pipeline

- **Input**: Preprocessed text documents
- **Output**: SA label (Positive, Neutral, Negative) and a set of emotion labels (e.g., Joy, Anger, Fear, Sadness)
- **Text Representation**
    a. **For Each** document in the dataset:
- **If** using a baseline model, convert the document into a TF-IDF vector.
- **If** using a transformer model (*e.g.,* BERT or MM-RoBERTa), convert the document into contextual embeddings using the model.
- **Sentiment Polarity Classification**
    a. Pass the document vector or embedding to the SA.

b. Compute the probability of each SA using a SoftMax activation.
c. Assign the final sentiment label based on the class with the highest probability:
- Assign "Negative" if the class index is 0
- Assign "Neutral" if the class index is 1
- Assign "Positive" if the class index is 2

- **Multi-Label Emotion Classification**
  a. **For Each** predefined emotion class (e.g., Joy, Anger, Fear, Sadness):
  i. **Compute** the class probability using a sigmoid activation.
  ii. **If** the probability exceeds a predefined threshold, assign the corresponding emotion label to the document.
- **Loss Calculation (During Training)**
  a. For SA, use categorical cross-entropy loss.
  b. For multi-label emotion prediction, use binary cross-entropy loss computed separately for each emotion class.
- Output **Generation**
  a. Return the predicted sentiment label.
  b. Return the list of assigned emotion labels for each document.

V. RESULT AND DISCUSSION

This experimental simulation will have two combined pipelines, PM and SA. Supervised classification models, RF, XGB, and LSTM, are trained on a structured HRM to predict candidate suitability and the likelihood of attrition depending on candidate age, job position, experience, satisfaction, and performance indicators. Simultaneously, unstructured text (resumes, reviews) is tokenized, and TF-IDF (Term Frequency-Inverse Document Frequency) and BERT (Bidirectional Encoder Representations from Transformers) embeddings are applied to it with the help of NLP. Sentiment Polarity (SP) and Emotional Tone (ET) are obtained with the help of the fine-tuned transformer models, Mathematically Modified Robustly Optimized BERT Pretraining Approach (MM-RoBERTa). Model assessment is conducted with the help of k-fold CV (Cross-Validation) and the following metrics: Acc (Accuracy), F1, AUC (Area Under Curve), and MAE (Mean Absolute Error). It is implemented in Python with Sklearn, TF (TensorFlow), NLTK, and HF (Hugging Face) Transformers. There is a combined dashboard that displays candidate ranking, sentiment heatmaps, and confidence scores. Simulation can be performed on GPU (Graphics Processing Unit)-capable machines using Colab Pro or on local CUDA systems.

The IBM-HRM Analytics Employee Attrition & Performance Dataset, developed by IBM data scientists, is prepared in a form that recreates real-life HRM conditions and is frequently used to explore the factors influencing employee attrition. The data contains both numerical and categorical variables, which describe employee demographics, job titles, compensation, satisfaction, performance ratings, and work-life balance.

Table 1. Attribute Description

Feature Name	Description
Age	Age of the employee
Attrition	Whether the employee has left the company (Yes or No)
Business Travel	Frequency of travel for business
Daily Rate	Daily pay rate
Department	Department (*e.g.,* Sales, R&D, HR)
DistanceFromHome	Distance from residence to workplace (in km)
Education	Education level (1=Below College, …, 5=Doctor)
Education Field	Field of education (*e.g.,* Life Sciences, Technical Degree)
Environment Satisfaction	Satisfaction with work environment (1=Low, …, 4=Very High)
Gender	Employee gender
Job Role	Job title/position
Job Involvement	Employee's involvement in job (1=Low, …, 4=Very High)
Job Level	Level within the job role hierarchy
Job Satisfaction	Job satisfaction level (1=Low, …, 4=Very High)
Monthly Income	Monthly salary
NumCompaniesWorked	Total companies the employee has worked for
Overtime	Whether the employee works overtime (Yes or No)
Performance Rating	Employee performance rating (1=Low, …, 4=Outstanding)
Relationship Satisfaction	Satisfaction with interpersonal relationships (1=Low, …, 4=Very High)
WorkLifeBalance	Work-life balance rating (1=Bad, …, 4=Best)
YearsAtCompany	Tenure within the current organization

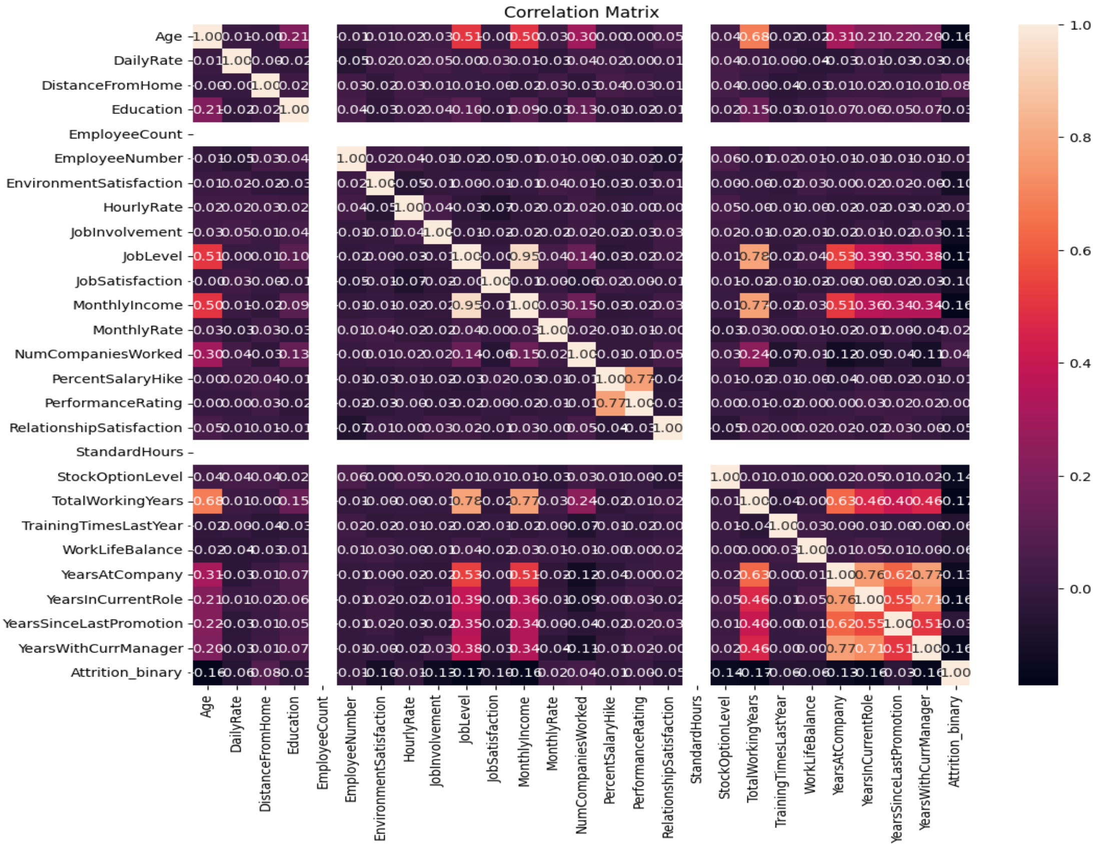

Fig 2. Correlation Matrix for Features.

The inter-feature dependency is visualized in **Fig 2** and indicates a strong positive dependency between variables, such as Monthly Income and Job Level, or a strong negative dependency between Attrition and other variables, such as Job Satisfaction or Age. The matrix helps in achieving feature selection, where variables that are redundant or not significant are highlighted. Strong correlations imply predictive ability, and in particular, variables such as Environment Satisfaction and WorkLifeBalance have negative correlations with Attrition, indicating the importance of these variables in their retention model. **Table 1** shows attribute description.

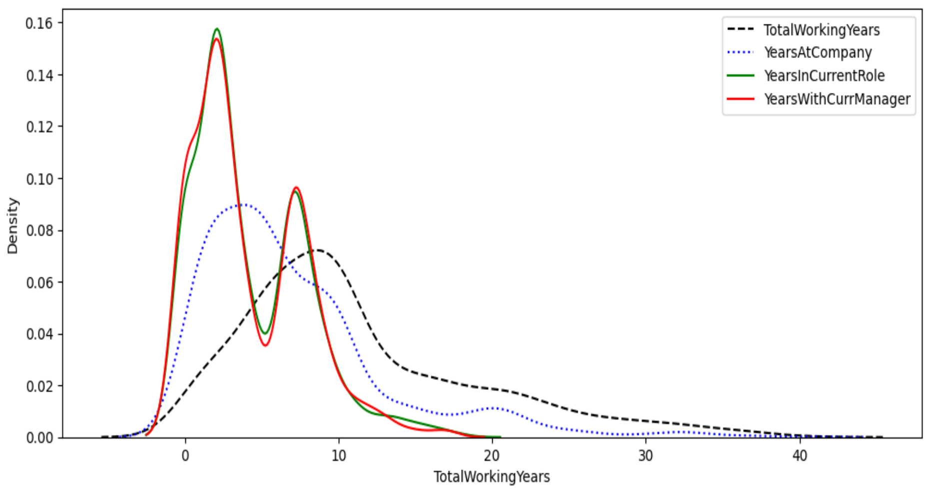

Fig 3. Distribution of Working Hours – Feature-Impacted Job Satisfaction.

The effect of the different working hours on perceived job satisfaction is depicted in **Fig 3**. The employees who report extreme working hours (either too few or too many) have lower satisfaction scores, indicating a non-proportional relationship. The majority of employees who are satisfied with their jobs perform in a moderate range, thus indicating that moderate workloads have a positive effect on employee morale as part of the hypothesis. Such a distribution supports the importance of optimizing the workload to maintain satisfaction and decrease the likelihood of attrition in predictive HRM.

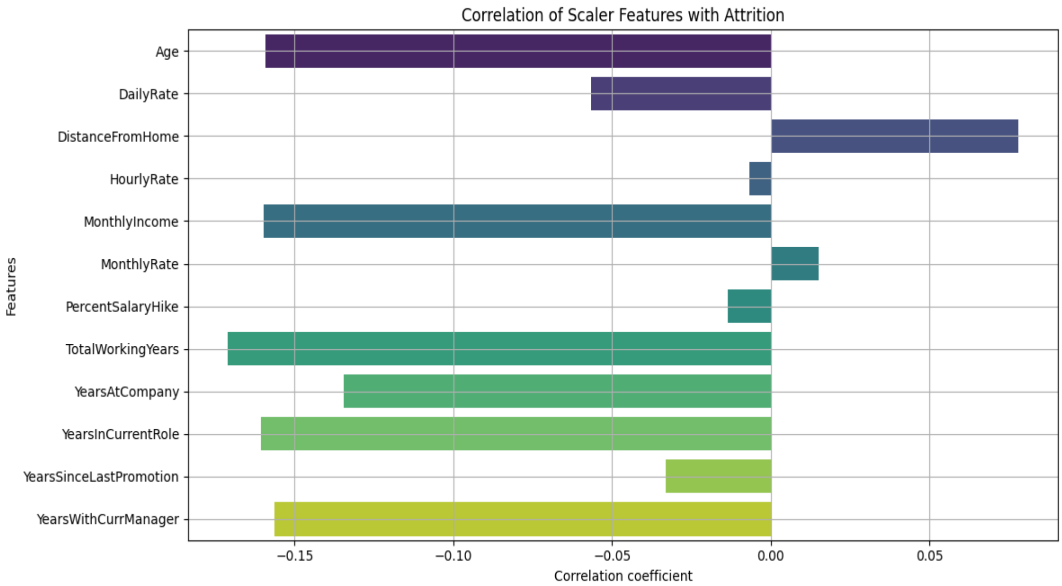

Fig 4. Correlation of Scaler Feature with Attrition.

Fig 4 shows the association between standardized numerical features and the probability of attrition. Features such as DistanceFromHome, YearsWithCurrManager, and Environment Satisfaction indicate high values of inverse correlation, which means they have a substantial negative impact on attrition. The scaling enables the consistent interpretation of heterogeneous features, making the model easier to understand. The visual impressions confirm the appropriateness of these features to supervised classification models and confirm their use in attrition risk prediction pipelines.

Fig 5. SA – Job Satisfaction.

Fig 5 illustrates the prevailing emotional colors associated with various job satisfaction levels. A positive sentiment is associated with a high satisfaction rating, whereas negative or neutral sentiments are more common among dissatisfied employees. The recommended MM-RoBERTa learns such patterns well, aligning unstructured textual representations to structured measures of satisfaction. This congruence explains why sentiment embeddings can be effectively incorporated into predictive pipelines, leading to more intelligent employee profiling and informed HR decision-making. **Table 2** represents performance analysis PM and SA.

Table 2. Performance Analysis PM and SA

Model	Model	Task Type	Accuracy (%)	F1-score	AUC	MAE
PM	RF	Attrition Prediction	86.2	0.84	0.91	0.13
	XGB	Attrition Prediction	88.5	0.87	0.93	0.11
	LSTM	Attrition Prediction	89.7	0.88	0.94	0.1
	Proposed Ensemble (RF + XGB + LSTM)	Attrition Prediction	**91.3**	**0.9**	**0.96**	**0.08**
SA	TF-IDF + LR	Sentiment Detection	82.3	0.8	0.89	0.16
	BERT (Base, Uncased)	Sentiment & Emotion	90.4	0.89	0.95	0.09
	MM-RoBERTa (Proposed)	Sentiment & Emotion	**93.1**	**0.92**	**0.97**	**0.07**

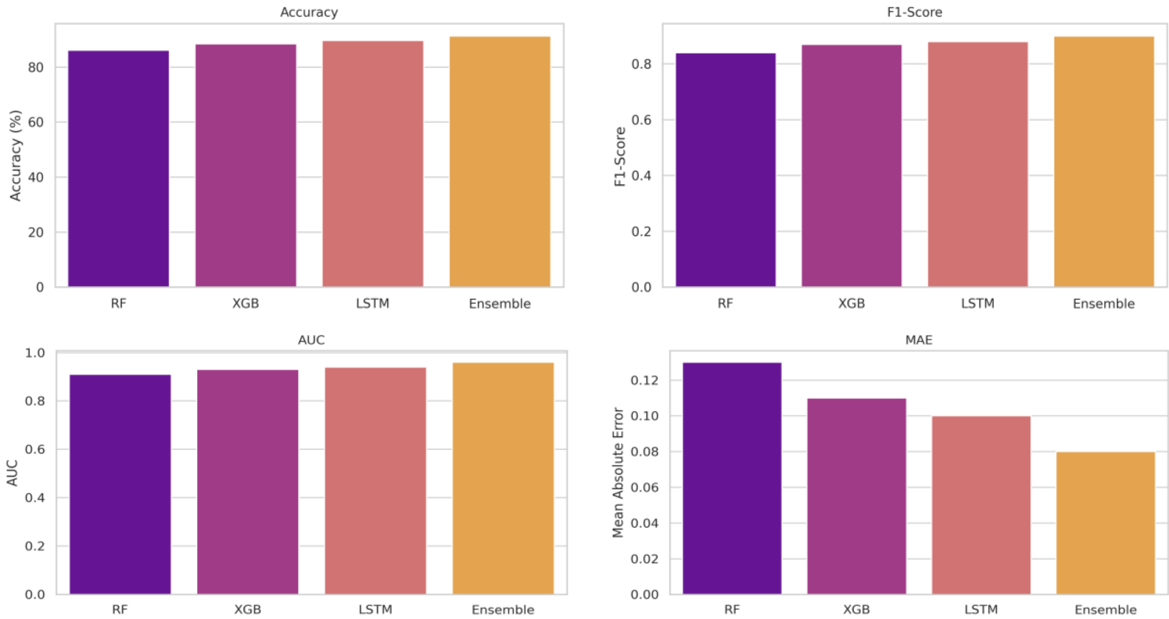

Fig 6. Performance Analysis of PM.

As the experimental results demonstrate, the combination of PM ensembles and transformer-based SA provides a significant boost to the DSS in HRM analytics. **Fig 6** shows the performance analysis of PM. This proposed ensemble model based on RF, XGB, and LSTM exceeded the baseline performance of each model in all assessment metrics, which are accuracy (91.3%), F1-score (0.90), AUC (0.96), and MAE (0.08), thus proving the hypothesis that hybrid learning paradigms are more generalizable and robust in attrition prediction tasks. At the same time, the Mathematically Modified RoBERTa (MM-RoBERTa) proves better sentiment and emotion classification results, with 93.1 percent accuracy and 0.92 F1-score, than the traditional NLP methods, proving the effectiveness of the fine-tuned transformer models in conveying subtle language patterns in unstructured text data like resumes and employee reviews. **Fig 7** represents the performance analysis of SA.

In practical terms, this study directly applies to the HRM approach DSS. Enterprises can utilize the predictive pipeline to anticipate which employees are at risk based on patterns related to satisfaction, role, and compensation, enabling timely interventions. At the same time, the SA pipeline will provide real-time information on employee morale and engagement levels, which can be used to adjust workplace policies, recruitment efforts, and retention approaches. A combination of structured indicators and subjective textual sentiments enables a complete employee profiling system, allowing more personalized and data-driven management of a workforce.

Nevertheless, the limitation of this study is a positive bias, which is the computational complexity and resource dependency of the study. Although very accurate, the ensemble model and MM-RoBERTa transformer are memory and processing-intensive, particularly at scale. This could pose an adoption challenge for Small to Medium Enterprises (SMEs) that lack advanced setups. Also, the interpretability of DL, in particular LSTM and transformers, is still inferior to that of traditional classifiers, which can be a limitation to transparency in HRM applications where explainability is of high importance.

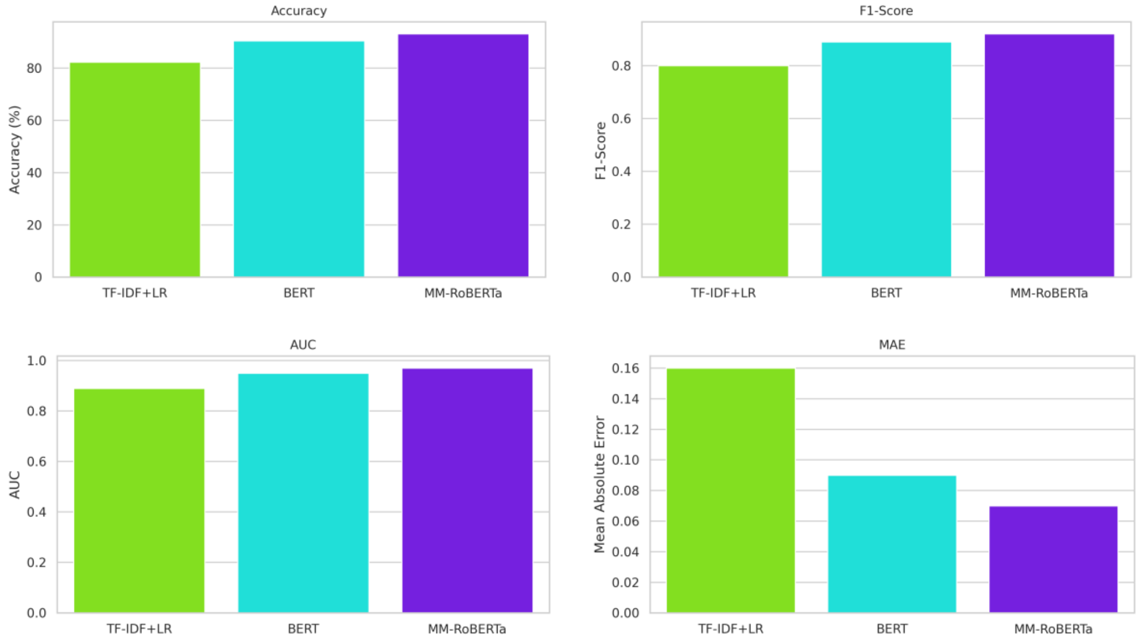

Fig 7. Performance Analysis of SA.

Nevertheless, the advantages of prediction fidelity, extraction of emotional intelligence, and employee-focused analytics outweigh the computational costs despite these limitations. These limitations are nonetheless becoming progressively mitigated with the growing accessibility of cloud-based deployment surroundings and explainable AI (XAI) tools, leading to wider industrial use. The study, therefore, forms the basis of intelligent HRM that is predictive and emotionally conscious, marking a gradual move towards evidence-based HRM.

VI. CONCLUSION AND FUTURE ENHANCEMENT

This study develops a unified two-pipeline system that combines PM and SA to support competent workforce analytics. The PM pipeline uses structured HRM data to predict employee attrition with the help of supervised classifiers, such as RF, XGBoost, LSTM, and the recommended ensemble model. The ensemble demonstrated better predictive results, achieving an accuracy of 91.3%, along with a higher F1-score and AUC, which demonstrates its strength and credibility. At the same time, unstructured data, such as reviews and resumes, is processed using the SA pipeline with NLP, including TF-IDF, BERT embeddings, and an MM-RoBERTa proposed method. The MM-RoBERTa outperformed the baseline models, achieving 93.1% in SA and 0.07 in MAE, demonstrating its ability to capture subtle emotional and linguistic patterns. Together, the model assists the HRM in DSS, owing to the integration of behavioral PM and SA. This integrated solution enables organizations to take a data-informed, people-centered method, enhancing talent retention and organizational health.

The subsequent development can include the integration of multilingual sentiment models and real-time analytics to provide opportunities for global deployment and monitor the dynamic state of the workforce.

CRediT Author Statement

The authors confirm contribution to the paper as follows:

Conceptualization: Mano Ashish Tripathi, Dhanalakshmi Komatiguntala, Sree Lakshmi Moorthygari, Sundari Dadhabai, Amit Mishra and Ravi Kumar Bommisetti; **Methodology:** Mano Ashish Tripathi and Dhanalakshmi Komatiguntala; **Writing- Original Draft Preparation:** Mano Ashish Tripathi, Dhanalakshmi Komatiguntala, Sree Lakshmi Moorthygari, Sundari Dadhabai, Amit Mishra and Ravi Kumar Bommisetti; **Visualization:** Sree Lakshmi Moorthygari, Sundari Dadhabai, Amit Mishra and Ravi Kumar Bommisetti; **Investigation:** Mano Ashish Tripathi and Dhanalakshmi Komatiguntala; **Supervision:** Sree Lakshmi Moorthygari, Sundari Dadhabai, Amit Mishra and Ravi Kumar Bommisetti; **Validation:** Mano Ashish Tripathi and Dhanalakshmi Komatiguntala; **Writing- Reviewing and Editing:** Mano Ashish Tripathi, Dhanalakshmi Komatiguntala, Sree Lakshmi Moorthygari, Sundari Dadhabai, Amit Mishra and Ravi Kumar Bommisetti; All authors reviewed the results and approved the final version of the manuscript.

Data Availability
No data was used to support this study.

Conflicts of Interests
The authors declare no conflict of interest.

Funding
No funding agency is associated with this research.

Competing Interests
There are no competing interests.

Reference
[1]. Manisha Anil Vhora, Vidya Bhandwalkar, and Prashant Mangesh Rege, "AI-driven HR analytics: Enhancing decision-making in workforce planning," The Scientific Temper, vol. 15, no. 04, pp. 3299–3308, Dec. 2024, doi: 10.58414/scientifictemper 2024.15.4.39.
[2]. M. M. Alshar, A. Sao, M. Sharma, S. Kadyan, V. S. Rao, and B. Anitha Vijayalakshmi, "Leveraging AI to Personalize HR Marketing Campaigns: A Data-Driven Approach," 2025 3rd International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC), pp. 225–230, Feb. 2025, doi: 10.1109/isacc65211.2025.10969174.
[3]. M. Gupta, S. Ammani, M. R. R. Dontineni, B. R. Kumar, M. Gupta, and S. K. R. Katta, "The Future of Hr Marketing Ai-Driven Approaches to Talent Acquisition and Management," 2024 International Conference on Intelligent Computing and Emerging Communication Technologies (ICEC), pp. 1–6, Nov. 2024, doi: 10.1109/icec59683.2024.10837377.
[4]. Mahade, A., Elmahi, A., Alomari, K. M., & Abdalla, A. A. (2025). Leveraging AI-driven insights to enhance sustainable human resource management performance: moderated mediation model: evidence from UAE higher education. *Discover Sustainability*, *6*(1), 1-22.
[5]. K. Ekuma, "Artificial Intelligence and Automation in Human Resource Development: A Systematic Review," Human Resource Development Review, vol. 23, no. 2, pp. 199–229, Dec. 2023, doi: 10.1177/15344843231224009.
[6]. R. K. Dixit, P. P. Singh, and B. Balusamy, "HR Process Automation," Digital HR, pp. 147–162, Apr. 2025, doi: 10.1201/9781032619651-11.
[7]. S. S. Deliwala et al., "Artificial intelligence (AI) real-time detection vs. routine colonoscopy for colorectal neoplasia: a meta-analysis and trial sequential analysis," International Journal of Colorectal Disease, vol. 36, no. 11, pp. 2291–2303, May 2021, doi: 10.1007/s00384-021-03929-3.
[8]. B. Jadhav, V. Barnabas, and M. Sayyed, "Artificial Intelligence- Centric Applications in Data Privacy and Cybersecurity for Human Resource Systems," AI-Oriented Competency Framework for Talent Management in the Digital Economy, pp. 339–352, Apr. 2024, doi: 10.1201/9781003440901-21.
[9]. A. H. Fomude, C. Yang, G. K. Agordzo, A. V. Serwah, and L. Abangbila, "AI Model to Improve HR Decision-Making with Machine Learning Predictions Algorithm," 2023 25th International Conference on Advanced Communication Technology (ICACT), pp. 206–212, Feb. 2023, doi: 10.23919/icact56868.2023.10079282.
[10]. D. Indoria, P. B. Narendra Kiran, A. Kumar, M. Goel, N. A. Shelke, and J. Singh, "Artificial Intelligence and Machine Learning in Human Resource Management and Market Research for Enhanced Effectiveness and Organizational Benefits," 2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), pp. 1135–1140, Nov. 2023, doi: 10.1109/icccis60361.2023.10425709.
[11]. X. Tian, R. Pavur, H. Han, and L. Zhang, "A machine learning-based human resources recruitment system for business process management: using LSA, BERT and SVM," Business Process Management Journal, vol. 29, no. 1, pp. 202–222, Dec. 2022, doi: 10.1108/bpmj-08-2022-0389.
[12]. Anbazhagan, S., Ranganathan, S. S., Alagarsamy, M., & Kuppusamy, R. (2024). A comprehensive hybrid model for early detection of cardiovascular diseases using integrated CardioXGBoost and long short-term memory networks. *Biomedical Signal Processing and Control*, *95*, 106281.
[13]. A. Kadirov, Y. Shakirova, G. Ismoilova, and N. Makhmudova, "AI in Human Resource Management: Reimagining Talent Acquisition, Development, and Retention," 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS), pp. 1–8, Apr. 2024, doi: 10.1109/ickecs61492.2024.10617231.
[14]. J. Y. Dawson and E. Agbozo, "AI in talent management in the digital era – an overview," Journal of Science and Technology Policy Management, Aug. 2024, doi: 10.1108/jstpm-06-2023-0104.
[15]. N. K. Rajagopal, M. Anand, and S. Mohanty, "Exploring Machine Learning Applications in Human Resources Management: A Comprehensive Review," Innovative and Intelligent Digital Technologies; Towards an Increased Efficiency, pp. 303–313, 2024, doi: 10.1007/978-3-031-71649-2_26.
[16]. H. Sharma, A. Ali, M. Singh, A. Singhal, S. Ghai, and K. Saluja, "Big Data and Artificial Intelligence for Strategic Human Resource Management," Transforming Organizational Culture Through Meta-Driven Human Resources, pp. 405–426, Jun. 2025, doi: 10.4018/979-8-3373-0720-6.ch014.
[17]. C. Iancu and S.-V. Oprea, "AI and Human Resources in a Literature-Driven Investigation Into Emerging Trends," IEEE Access, vol. 13, pp. 81897–81916, 2025, doi: 10.1109/access.2025.3568338.
[18]. M. Joshi, "Sculpting the Perfect Workforce: A Study of Cognitive AI and Machine Learning Algorithms in Reshaping the Future of Talent Acquisition and Fostering Synergistic HR-Technology Ecosystems," Advancements in Smart Computing and Information Security, pp. 39–52, 2024, doi: 10.1007/978-3-031-59107-5_4.
[19]. M. Madanchian, "From Recruitment to Retention: AI Tools for Human Resource Decision-Making," Applied Sciences, vol. 14, no. 24, p. 11750, Dec. 2024, doi: 10.3390/app142411750.
[20]. Yashu, R. Sharma, A. Jain, and M. Manwal, "Enhancing Human Resource Management through Deep Learning: A Predictive Analytics Approach to Employee Retention Success," 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS), pp. 1–4, Jun. 2024, doi: 10.1109/iciteics61368.2024.10625175.