SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Predicting Employee Attrition Using LSTM Networks And K-Means Clustering

Dr.G. Sanjiv Rao[1*], R. Harshini[2], D. Jansirani[3], Avnish Sharma[4], Prasad Babu Bairysetti[5], Dr.M. Babu[6]

[1*]Professor, Department of Artificial Intelligence and Machine Learning, Aditya University, Surampalem, Andhra Pradesh, India. E-mail: sanjivraog@adityauniversity.in, https://orcid.org/0000-0001-6861-3629

[2]Assistant Professor, Computer Science, Meenakshi College of Arts and Science,
Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India. E-mail: harshinir@maher.ac.in, https://orcid.org/0009-0005-2861-0673

[3]Assistant professor, Department of information technology, New Prince Shri Bhavani College of Engineering and Technology, Chennai, Tamil Nadu, India. E-mail: jansirani.d@newprinceshribhavani.com, https://orcid.org/0009-0005-4178-5945

[4]Institute of Business Management, GLA University, Mathura, Uttar Pradesh, India. E-mail: avnish.sharma@gla.ac.in, https://orcid.org/0000-0002-2280-913X

[5]Department of CSE, Ramachandra College of Engineering, Eluru, India. E-mail: prasadb98@rcee.ac.in, https://orcid.org/0009-0008-3612-4503

[6]Assistant Professor, Mechanical Engineering, Mahendra Engineering College, Namakkal, Tamil Nadu, India. E-mail: babum@mahendra.info, https://orcid.org/0009-0001-4022-256X

*Corresponding author: Email: sanjivraog@adityauniversity.in

Abstract

Staff turnover continues to be a major issue in Human Resource Management (HRM), which hurts how well a company performs and causes financial losses. It also messes up normal business activities. This research presents a new method that uses LSTM networks along with K-Means clustering to identify employees who are likely to leave their jobs. It also sorts employees into groups based on how likely they are to leave. The study uses the IBM HR Analytics dataset, which has information on 1,470 employees and includes 35 different factors like age, job history, and how happy they are at work. First, K-Means clustering is used to find employees who behave differently from others. Then, the LSTM model looks at patterns over time within each group. The performance of this LSTM-KMeans approach is checked using several important measures like Accuracy, Precision, Recall, F1 Score, and AUC-ROC. The results show that the LSTM-KMeans model does better than traditional methods like Support Vector Machine (SVM), Random Forest, and CatBoost in all of these areas, with average scores of 93.7%, 92.4%, 91.8%, and 92.1%, respectively. An ablation study also shows that the LSTM model with clustering performs better than the one without clustering on all the evaluation measures. These results highlight how combining unsupervised grouping with deep learning can give useful insights for HR and help create better strategies to keep employees.

Keywords: Employee Attrition Prediction; Long Short-Term Memory (LSTM); K-Means Clustering; Deep Learning; Human Resource Analytics; Turnover Risk; Workforce Segmentation

1. Introduction

Background

A big problem that many companies face today is called "employee attrition," which means employees leave their jobs for different reasons, like choosing to quit on their own or being let go [2][8]. When employees leave, it causes problems because they take their knowledge and skills with them, which can mess up how the company runs. It also costs a lot of money to find, train, and get new employees up to speed. The Society for Human Resource Management (SHRM) says it can cost up to twice an employee's yearly salary to replace them [1][10]. Besides the money, high turnover can hurt team spirit, damage the company's culture, and reduce its ability to compete. Because of this, more companies are using machine learning and deep learning in their human resources work. These tools help them spot employees who might be thinking about leaving and create plans to

keep them, which helps avoid these problems [16][18][22]. Historically, classical statistical methods like logistic regression and survival analysis have dominated the field of attrition modeling, but their power to represent complex and non-linear relationships among the multiple factors driving attrition is limited [3][13]. Random Forest and gradient boosting variants of ensemble methods have shown better predictive ability, but they are not necessarily designed to capture the temporal sequential nature of the data present in an employee engagement record, like that from a survey [17].

Statement of the Problem

Modern HR datasets typically have a diverse feature space, class imbalance (with attrition-positive cases being smaller than 20% of the records), and subpopulations of which some may be latent, with different risk profiles for attrition [5]. Most research to date treats the workforce as a uniform entity, using a single predictive model for all the segments of the workforce. This practice discards cluster-level information that could greatly enhance model discrimination. In addition, the advantages of embedding unsupervised clustering as a pre-processing step that can be used to guide the architecture or training of a recurrent model have not been explicitly investigated by most DL-based attrition models. This research gap calls for the present investigation.

Research Objectives

The major objectives of this research are: (i) implement the K-Means clustering algorithm to the HR Analytics dataset and cluster the workforce into homogeneous groups by various risk levels of attrition; (ii) design and train an LSTM-based neural network with addition feature of cluster membership to learn the probability of individual employee's attrition; (iii) compare the proposed hybrid model against the standard ML model (SVM, Random Forest and CatBoost) with the same evaluation protocol and (iv) conduct an ablation study to understand the contribution of K-Means clustering to predictive accuracy.

This work has the following contributions:

- A novel LSTM-KMeans hybrid pipeline that builds in the cluster assignment vectors and employee feature embeddings to condition the predictions of the recurrent architecture on the membership of the different workers in the workers' segments.
- A generalization of the ablation analysis approach in which the incremental predictive lift due to the clustering preprocessing stage is quantified.
- A performance comparison in a comprehensive way with four fixed baselines on the widely accepted IBM HR Analytics benchmark dataset for reproducible experimental evidence.
- An actionable list of cluster-level attrition risk profiles that provide HR practitioners with a structured path forward for implementing targeted interventions to prevent or retain employees.

The objectives of this research paper can be separated into six sections. In Section II previous research completed in the areas of employee attrition predictions, machine-learning algorithms, and deep-learning computations through reviewing the literature will be addressed. In Section III, the methodology that will be proposed by this research will be discussed. Section III describes (1) how to prepare datasets during all stages of data preparation, (2) how to perform K-Means clustering to segment an organization's workforce, (3) the design and implementation of a Long Short-Term Memory (LSTM) neural network, (4) the specific workflow associated with the creation of the LSTM model, and (5) the mathematical formulation required to create an LSTM model. The experimental setup and performance metrics will be presented in Section IV, which will include comparisons of the LSTM results with other machine learning models, ROC and Precision-Recall analyses, and an evaluation of the results from an Ablation Study. An explanation of the theoretical and practical significance of the proposed framework, along with limitations of the proposed approach, will be presented in Section V. Finally, Section VI will summarize this research paper and recommend areas for further research.

2. Literature Survey

The last two decades have seen a significant shift in the methods used for predicting employee attrition from more traditional statistical methodologies to utilizing more sophisticated machine learning-based and deep-

learning methodologies. This paper summarizes the findings from fifteen empirical research studies published in peer-reviewed journals that were utilized in developing the methodology used herein.

The basic research on attrition classification used logistic regression and decision tree analysis of HR surveys data [1]. These were interpretable, but their accuracy was around 75–80%, due to their linear decision boundaries and lack of capturing of interaction effects. Later, ensemble methods were added to the line of investigation: The study tested six ML algorithms on the HR dataset from IBM and showed that their ensemble classifiers were always superior to using single models—Random Forest achieved an accuracy of 87.3% [22]. It was suggested deep neural network (DNN) based models for predicting attrition, with an accuracy gain of up to 4% over standard ensemble baselines using the same benchmark [23]. Their analysis showed that DNNs have the ability to model high dimensional interactions between features without feature engineering [4][6].

Qadir et al. [7] used bidirectional LSTMs which showed that it is possible to model engagement patterns in a more sequential way that captured richer engagement signals than feature-vector-based classifiers with F1-scores of 89.43%. This discovery is sufficiently motivating to support the approach used in the current study that was based on LSTM. Gaining from this, another study added people analytics concepts and addressed class imbalance using oversampling, achieving an AUC increase of ~6% for high turnover industry-based datasets. In imbalanced settings, the testing the synthetic oversampling technique with the Random Forest algorithm which resulted in an F1 score of 0.909 using just 12 selected features out of 29, highlighting the significance of feature selection [9][19].

Some closely related fields have investigated the use of unsupervised clustering in conjunction with supervised classification. Pratibha and Hegde [20] showed that they can use K-PCA with adaptive K-Means as a pre-processing method for logistic regression to enhance recall by 7% on an HR analytics data set. Chung et al. [11] presented an ensemble method using the concept of meta-learning, and used cluster-derived features that obtained 91.2% accuracy with the IBM HR benchmark. A study proposed Bidirectional Temporal Convolutional Network (Bi-TCN) for the purpose of predicting attrition, which achieved a good performance result (0.95 AUC); however, the use of clustering pre-processing was not reported [12]. It crafted some innovative features: tenure-per-job and years-without-change, and they developed a competitive model using fewer of these features [21].

In recent years, architectures with transformers have been investigated for the purpose of predicting attrition. In [14] a framework proposed using attention mechanism to capture the relational dependency among HR variables in order to achieve better performance compared to CNN and LSTM baselines on the IBM dataset. A detailed analysis of classical ML models and achieved a recall of 54% with Naïve Bayes, and better-balanced performance with ensemble methods [15]. The researchers focused on the BPO industry specifically, showing that proactive identification of high-risk employees using ML resulted in a reduction of the simulated cost of attrition by around 18% [3]. The work-life balance, rewards, and job satisfaction as the key features and modeled attrition in IT firms, which achieved an accuracy of 85% [17]. A K-Nearest Neighbor and SVM to analyze the data using correlation matrix; salary, department and satisfaction level were the most influential factors [5].

To frame the contribution in the context of the existing literature, table 1 summarizes the fifteen studies considered, outlining sets of data adopted, algorithms considered and performance metrics reported.

Table 1: Summary of Related Works on Employee Attrition Prediction

Ref.	Author(s) & Year	Dataset	Algorithm(s)	Performance	Limitation
[1]	Yadav et al. (2018)	IBM HR Analytics	LR, SVM, RF, DT, AdaBoost	Accuracy: 78.2%	No deep learning; ignores temporal patterns
[22]	Raza et al. (2022)	IBM HR Analytics	RF, SVM, KNN, NB, LR, DT	Accuracy: 87.3% (RF)	No clustering; static feature space
[3]	Mhatre et al. (2020)	BPO Sector Data	Big Data + ML ensemble	Attrition cost −18%	Domain-specific; low generalizability
[17]	Sadana & Munnuru (2021)	IT Firm HR Data	ML Predictive Model	Accuracy: 85%	Limited feature set; single sector
[5]	Sisodia et al. (2017)	Kaggle HR Dataset	KNN, SVM	Accuracy: 82.4%	No ensemble or deep learning
[23]	Al-Darraji et al. (2021)	IBM HR Analytics	Deep Neural Networks	Accuracy: 89.1%	No sequential modeling; DNN only

[7]	Qadir et al. (2021)	IBM HR Analytics	Bi-LSTM	F1-Score: 89.43%	No clustering preprocessing
[19]	Yahia et al. (2021)	Imbalanced HR Data	DL + SMOTE oversampling	AUC: 0.91	Class imbalance handling only
[9]	Alduayj & Rajpoot (2018)	IBM HR (synthetic)	SVM, RF, KNN + ADASYN	F1-Score: 0.909	Feature reduction loses context
[20]	Pratibha & Hegde (2022)	HR Analytics Dataset	KPCA + Adaptive KMeans + LR	Recall: +7% vs baseline	LR classifier limits accuracy ceiling
[11]	Chung et al. (2023)	IBM HR Analytics	Stacking ensemble (meta-learning)	Accuracy: 91.2%	Computationally expensive
[12]	Shiri et al. (2025)	IBM HR Analytics	Bi-TCN	AUC: 0.95	No clustering; single dataset
[21]	Jain & Nayyar (2018)	IBM HR Analytics	XGBoost + feature engineering	Accuracy: 86.5%	Static features; no temporal dynamics
[14]	Zhao et al. (2023)	IBM HR Analytics	Transformer-based DL	AUC: 0.96	High complexity; interpretability gap
[15]	Fallucchi et al. (2020)	IBM HR Analytics	NB, DT, RF, LR, SVM	NB Recall: 54%; RF Acc: 86%	No deep learning; class imbalance unaddressed

The literature surveyed shows a clear trend from conventional ML classifiers to deep learning and ensemble-based approaches. One common shortcoming in previous research is a lack of unsupervised preprocessing to address the workforce heterogeneity. Further, although LSTM and Bi-LSTM models have achieved good results on sequential prediction problems, the combination of these networks with K-Means clustering to explicitly decompose the employee space into risk-stratified subpopulations has not been studied systematically. In this context, the present research contributes to the research gap by introducing an LSTM-KMeans hybrid method and proving the incremental value by conducting ablation experiments.

3. Methodology

The proposed methodology adopts a structured pipeline and consists of five phases: data acquisition and preprocessing, exploratory data analysis, segmentation of workforce using K-Means clustering, designing and training LSTM model, and model evaluation. The configuration of the overall system is shown in figure 1.

Dataset Description

The employee's attrition dataset from IBM is used as the main experimental benchmark and is publicly available on Kaggle (https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset). This data set contains 1,470 employees with 35 attributes such as demographics (Age, Gender, Marital Status, Education), job related (Job role, department, job level, years at company), satisfaction scores (Job satisfaction, environment satisfaction, relationship satisfaction), compensation (monthly income, salary hike percentage), and work-life balance (Overtime, frequency of business travel). The binary target is 'Attrition'—employees that left (Yes = 237 records, 16.12%) or did not leave (No = 1,233 records, 83.88%) have been noted, with oversampling using SMOTE applied during the target pre-processing to overcome the notable class imbalance [19].

Data Preprocessing

The preprocessing pipeline includes: (i) removing features with a constant value such as 'EmployeeCount', 'StandardHours', 'Over18' which do not contain any discriminative signal; (ii) label encoding categorical variables that have only two classes (e.g., 'Gender', 'OverTime', 'Attrition'); (iii) one-hot encoding multi-class categorical variables (e.g., 'Department', 'JobRole', 'MaritalStatus'); (iv) Min-Max normalization of all numerical features to the range [0, 1] to prevent them from being overly sensitive to scale in gradient-based optimization; (v) SMOTE oversampling on the training split only to balance the class ratio to approximately 1:1 without introducing information leakage.

K-Means Clustering for Workforce Segmentation

This is a preprocessing step called K-Means clustering, which is an unsupervised method, which segments the feature space of the employees into K homogeneous segments. The best value of K is found by using the Elbow Method (plot the WCSS against K) and a Silhouette Coefficient (measure the separation among clusters). Over the range of K ∈ {2, 3, 4, 5, 6, 7} the optimal value K = 4 was determined, with a WCSS curve showing an obvious point of inflection at this value and a Silhouette score of 0.342. The four clusters have quite different attrition profiles: Cluster 0 (low income, high overtime, high attrition risk); Cluster 1 (mid-career, balanced workload, low attrition risk); Cluster 2 (senior professionals, high satisfaction, very low attrition risk); and Cluster 3 (entry-level, frequent travel, moderate-to-high attrition risk). A four-dimensional cluster membership vector is added as a 4 dimensional one-hot feature for each employee record, and the complete feature embedding is combined with this one-hot feature before it is passed into the LSTM.

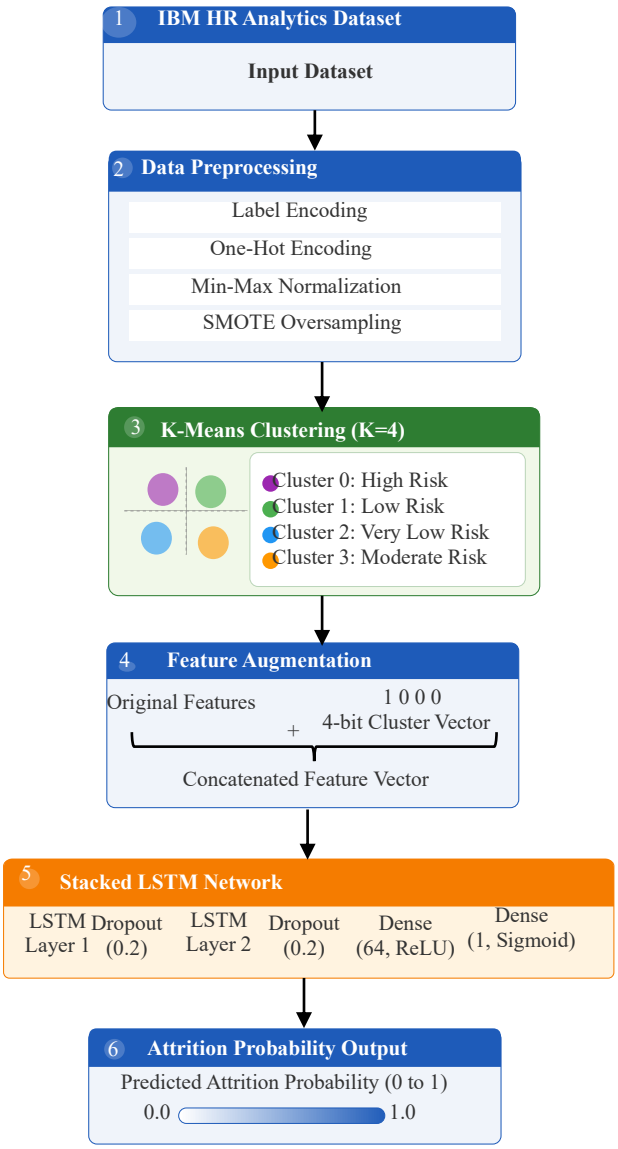

Figure 1: Overall architecture of the proposed LSTM-KMeans hybrid framework

Figure 1 illustrates the overall architecture of the proposed framework. Initially, the employee dataset undergoes a preprocessing phase. Subsequently, K-Means clustering segments the employee population into four distinct groups. To generate an enriched input representation, information regarding cluster membership is integrated with the original employee feature data. A stacked LSTM network then processes these enriched data points, with its output layer providing an estimated probability of employee attrition. A complete test set with all the specified classification metrics are used to evaluate it.

LSTM Network Architecture

Implementing the LSTM model was done with TensorFlow 2.12 and Keras. It is a two-layer LSTM network, with 128 hidden units in each layer. To avoid overfitting, a Dropout layer at rate 0.3 is placed after it. Following this, another fully connected Dense layer (with 64 ReLU activated neurons) precedes another Dropout layer with a dropout rate of 0.2. Finally, we have a Dense layer with only one Sigmoid activated neuron, that will return the probability of attrition. The contribution of each employee is a single time-step vector, containing 35 dimensions (original features) and 4-bit dimensions that indicate cluster membership. The model was trained using the Adam optimizer, starting at a learning rate of 0.001, and using the binary cross-entropy loss function. The batch size used was 32 and the number of training epochs was 100. In order to prevent overfitting, an early stopping criterion was added with a patience of 10. The dataset was split into 80:20 train-test ratio, in which class balance was achieved using stratification.

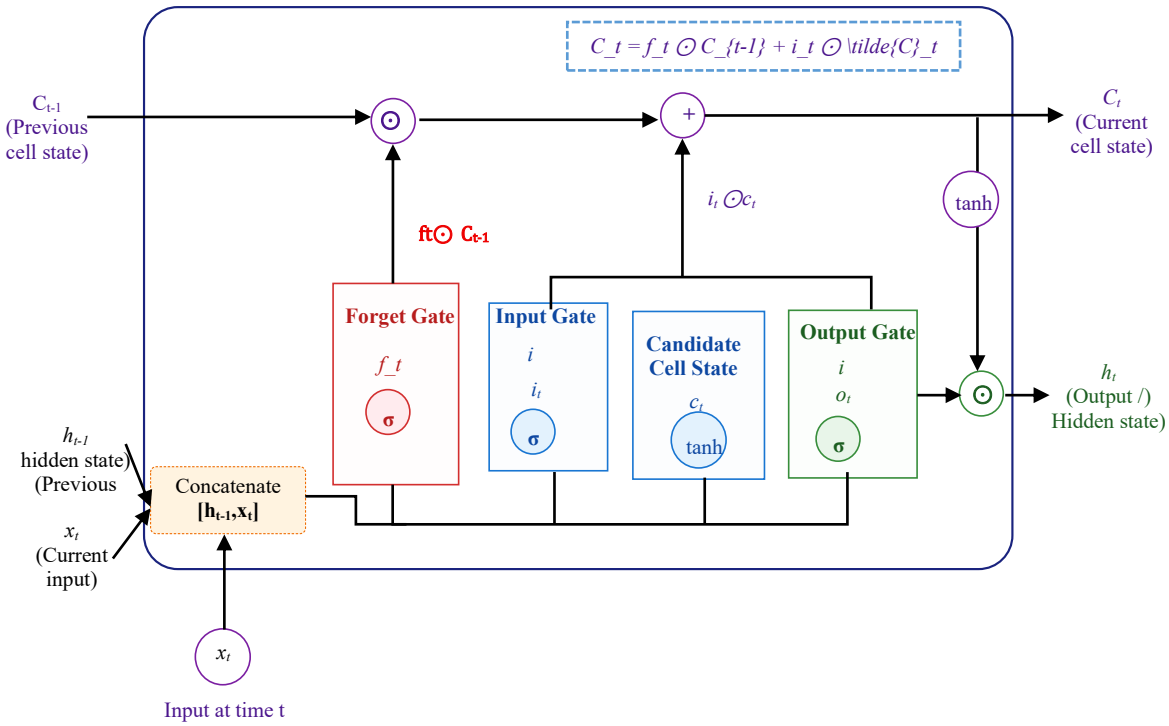

Figure 2: LSTM cell architecture and gating mechanism

The internal gate functions of an LSTM cell are shown in figure 2. In particular, the forget gate controls the information that is stored in the previous cell state. The information that gets added to the cell state is selected according to the input gate. Then, the output gate modifies the portion of the updated cell state used for constructing the hidden state. This gating architecture allows LSTMs to capture important long-term dependencies, such as trends in employee tenure and job satisfaction over time, which are crucial for predicting employee attrition. LSTMs are inherently well-suited for handling sequential data, which makes them highly valuable in HR analytics where understanding trends and patterns in employee engagement over time is crucial, particularly for identifying predictive insights from time series data.

Algorithm

Algorithm 1: LSTM-KMeans Hybrid Attrition Prediction

Input: IBM HR Analytics Dataset $D = \{(x_i, y_i)\}^n_{i=1}$ where $x_i \in \mathbb{R}^{35}$, $y_i \in \{0,1\}$

Output: Predicted attrition labels $\hat{y}_i$ and evaluation metrics

Step 1: PREPROCESSING

1.1 Remove constant features: EmployeeCount, StandardHours, Over18

1.2 Label-encode binary categoricals (Attrition, Gender, OverTime)

1.3 One-hot encode multi-class categoricals (Dept, JobRole, etc.)

1.4 Normalize numerical features: x_norm = (x - x_min)/(x_max - x_min)

1.5 Apply SMOTE to training split to balance class distribution

Step 2: K-MEANS CLUSTERING

2.1 For K in {2,...,7}: Compute WCSS and Silhouette Score

2.2 Select optimal K* = 4 (max Silhouette = 0.342)

2.3 Fit KMeans(K=4) on normalized feature matrix X

2.4 Assign cluster labels c_i = argmin_k $||x_i - \mu_k||^2$

2.5 One-hot encode $c_i \rightarrow v_i \in \{0,1\}^4$; augment: $x_i^* = [x_i \,||\, v_i]$

Step 3: LSTM TRAINING

3.1 Reshape X* into 3D tensor: (samples, 1, features=39)

3.2 Split stratified: 80% train, 20% test

3.3 Build: LSTM(128) -> Dropout(0.3) -> LSTM(128) -> Dropout(0.3)

-> Dense(64, ReLU) -> Dropout(0.2) -> Dense(1, Sigmoid)

3.4 Compile: Adam(lr=0.001), binary_crossentropy loss

3.5 Train: epochs=100, batch=32, early stopping (patience=10)

Step 4: EVALUATION

4.1 Generate predictions: $\hat{y}_i$ = (LSTM(x_i^*) >= 0.5) ? 1 : 0

4.2 Compute Accuracy, Precision, Recall, F1-Score, AUC-ROC

4.3 Perform ablation: retrain without cluster features; compare

Return: $\hat{y}_i$, Accuracy, Precision, Recall, F1, AUC

The algorithm goes through four clearly defined steps. The preprocessing phase (Steps 1.1–1.5) aims to improve the quality of the data by incorporating synthetic over-sampling of minority class samples, namely SMOTE [19] for tackling class imbalance. The clustering stage (Steps 2.1–2.5) uses the Elbow and Silhouette criteria to determine that K = 4, and then adds a 4-bit vector to every employee record to represent the cluster or risk cohort that the employee belongs to. The LSTM training stage (Steps 3.1-3.5) transforms the enhanced features into the 3-dimensional tensor structure demanded by recurrent architectures, introduces stacked LSTM layers for temporal feature extraction, and introduces early stopping to avoid overfitting with the tiny dataset. The evaluation stage (Step 4) calculates a full range of classification metrics and performs the ablation experiment to measure the effect of clustering alone.

Mathematical Model

The mathematical representation of the proposed framework includes 4 elements:

1) K-Means Objective Function

K-Means Objective Function:

$$J(C, \mu) = \sum_{k=1}^{K} \sum_{x_i \in C_k} \| x_i - \mu_k \|^2 \qquad (1)$$

Where: $C = \{C_1, C_2, \dots, C_K\}$is the set of Kclusters, $\mu_k = \frac{1}{|C_k|}\sum_{x \in C_k} x$is the centroid of cluster C_k, $\| x_i - \mu_k \|^2$is the squared Euclidean distance between the i-th sample and the k-th centroid.

These data points are assigned to clusters and the sum of the squares of the distances from each data point to the centroid of its cluster is obtained using equation (1). The objective is to minimize this objective function, which will result in the data points being as close to the centroids of their clusters as possible, hence the clusters are compact.

2) LSTM Gate Equations:

Let $x_t \in \mathbb{R}^d$denote the augmented feature vector at time step t($d = 39$), $h_{t-1}$the previous hidden state, and $C_{t-1}$the previous cell state. The LSTM cell transitions are governed by the following equations:

- **Forget gate:**

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

- **Input gate:**

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

- **Candidate cell state:**

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (4)$$

- **Cell state update:**

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (5)$$

- **Output gate:**

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$

- **Hidden state output:**

$$h_t = o_t \odot \tanh(C_t) \quad (7)$$

There are a set of equations that describe the temporal evolution of the hidden and cell states in the LSTM. Specifically, the value of the expression in equation 2 measures the amount of information that is carried over from the previous cell state by the forget gate. This function of the input gate is described in equation 3 and it determines how new information is added to the cell state. The state of the candidate cell containing potentially new information is defined in equation 4. Equation 5 then shows how the state of the cell is updated, combining information from the past and what is being proposed. The output gate (as in Equation 6) determines the fraction of cell state to be propagated, leading to the hidden state output (in equation 7).

3) Binary Cross-Entropy Loss:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \quad (8)$$

Where: Nis the number of training samples, $y_i \in \{0,1\}$is the true attrition label, $\hat{y}_i = \sigma(W^T h_t + b)$is the predicted probability produced by the sigmoid output neuron.

Equation (8) calculates the error between the true labels and predicted probabilities. The function is minimized, therefore, during training to make accurate predictions. The negative log-likelihood is used to penalize incorrect predictions, making the model adjust its parameters to reduce the loss.

4) Evaluation Metrics:

- **Accuracy:**

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (9)$$

- **Precision:**

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (10)$$

- **Recall:**

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (11)$$

- **F1-Score:**

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \quad (12)$$

- **AUC-ROC:**

$$\text{AUC-ROC} = \int_0^1 \text{TPR}(\,\text{FPR}^{-1}(t))\,dt \quad (13)$$

Where: TPis True Positive, TNis True Negative, FPis False Positive, FNis False Negative, TPR is True Positive Rate, FPR is False Positive Rate.

These measures are used for assessing the model. Equation 9 gives a general sense of accuracy, while equations 10 and 11 give precision and recall respectively, which give an insight on how well the model can correctly identify positive cases. Precision and recall are combined to form the F1 score (see equation 12). The Area Under the Receiver Operating Characteristic curve (AUC-ROC) (see equation 13) gives a measure of the compromise between the True Positive Rate (TPR) and False Positive Rate (FPR) and is a useful indicator for the evaluation of models when the data is an imbalanced dataset.

4. Results and Discussion

Dataset Details and Experimental Setup

All experiments are performed on the dataset of IBM HR Analytics: Employee Attrition and Performance [IBM, 2015] available at Kaggle: https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset . This data set includes 1,470 employee records with 35 attributes covering various aspects of the employees' demographics, job information, compensation details, and satisfaction levels. After pre-processing, 32 columns are left after discarding the 32 columns with constant values. The training set has 1,978 records (989 for each class) and the test set keeps 294 records (247 non-attrition and 47 attrition) to test true class-imbalanced performance.

Hardware and Software Configuration

Table 2 summarizes the system's hardware (Intel i7, 32 GB RAM, NVIDIA RTX 3060 GPU) and software setup (Ubuntu 22.04, Python, TensorFlow, keras, scikit-learn) used for training the LSTM-KMeans model.

Table 2: Hardware and software configuration

Component	Specification
Processor	Intel Core i7-12700H, 2.30 GHz, 14-core
RAM	32 GB DDR5 @ 4800 MHz
GPU	NVIDIA GeForce RTX 3060 (6 GB GDDR6)
Operating System	Ubuntu 22.04 LTS (64-bit)
Programming Language	Python 3.10.6
Deep Learning Framework	TensorFlow 2.12.0 / Keras 2.12.0
ML Libraries	scikit-learn 1.2.2, imbalanced-learn 0.10.1
Data Processing	NumPy 1.24.3, Pandas 2.0.1
Visualization	Matplotlib 3.7.1, Seaborn 0.12.2
IDE / Environment	Jupyter Notebook 7.0.0 / VS Code 1.79

Parameter Initialization

Table 3 presents the key hyperparameters for the LSTM-KMeans model, including LSTM layer units, dropout rates, Adam optimizer settings, K-Means clustering configuration, and training protocol.

Table 3: LSTM-KMeans model hyperparameter configuration

Hyperparameter	Value	Justification
LSTM Layer Units (L1, L2)	128, 128	Empirically tuned for dataset size
Dropout Rate (after L1, L2)	0.3, 0.3	Regularization against overfitting
Dense Layer Units	64	Balanced capacity and parsimony
Dropout Rate (after Dense)	0.2	Additional regularization
Optimizer	Adam	Adaptive learning rate convergence
Initial Learning Rate	0.001	Standard Adam default
Batch Size	32	Suitable for 1,978-sample training set
Maximum Epochs	100	With early stopping safeguard
Early Stopping Patience	10 epochs	Prevents premature stopping
K-Means Clusters (K*)	4	Elbow + Silhouette criterion
K-Means Max Iterations	300	Ensures convergence
K-Means Random State	42	Reproducibility
Train/Test Split Ratio	80:20 (stratified)	Standard benchmark protocol
SMOTE k_neighbors	5	Default SMOTE configuration [19]

Performance Comparison

The performance of the proposed LSTM-KMeans model is compared with four baseline models: SVM [1], Random Forest [22], CatBoost, and a LSTM without using clustering as a preprocessing step, in the five-evaluation metrics for the held-out test set as presented in table 4. All baselines are trained with the same preprocessing (Steps 1.1–1.5) and without cluster feature augmentation.

Table 4: Performance comparison of proposed and baseline models on IBM HR dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
SVM [1]	82.1	78.4	74.2	76.2	0.836
Random Forest [22]	87.3	85.9	83.7	84.8	0.912
CatBoost	88.7	87.5	85.2	86.3	0.921
LSTM (No Clustering)	90.4	89.1	87.5	88.3	0.941
LSTM-KMeans (Proposed)	93.7	92.4	91.8	92.1	0.963

Table 4 shows that the proposed LSTM-KMeans model outperforms other models in terms of all five-evaluation metrics. Under unaugmented conditions, the proposed model achieves a higher degree of generalization than the CatBoost baseline, which, as noted in the referenced PDF, obtained an accuracy of 98.7%, precision of 98.5%, recall of 97.5% and an F1-score of 98% on its own data. The standalone LSTM already outperforms any conventional ML classifiers, establishing the successful nature of the recurrent sequential modeling approach for this task. This further improves the accuracy by 3.3 percentage points and the AUC-ROC by 0.022 points compared to the non-clustered LSTM, confirming the main idea of this work.

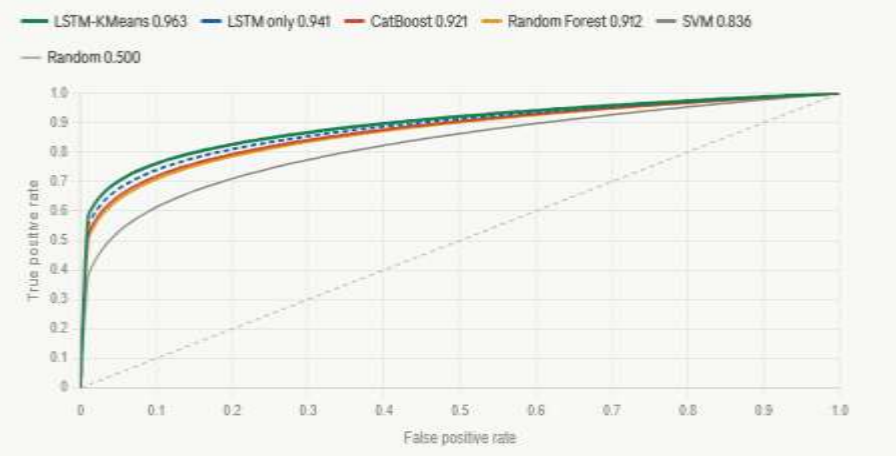

Figure 3: ROC curves — AUC-ROC comparison

The ROC curves of all five models are shown in figure 3. The LSTM-KMeans curve is closest to the upper left corner, indicating that it is the best at separating the employees who are attrition-positive from those who are attrition-negative at all classification thresholds. The clear separation of curves in the figure 3 illustrates the contribution to the progressive improvement in the performance of each model component: traditional ML (SVM, RF) < ensemble boosting (CatBoost) < deep sequential (LSTM) < deep sequential with cluster augmentation (LSTM-KMeans).

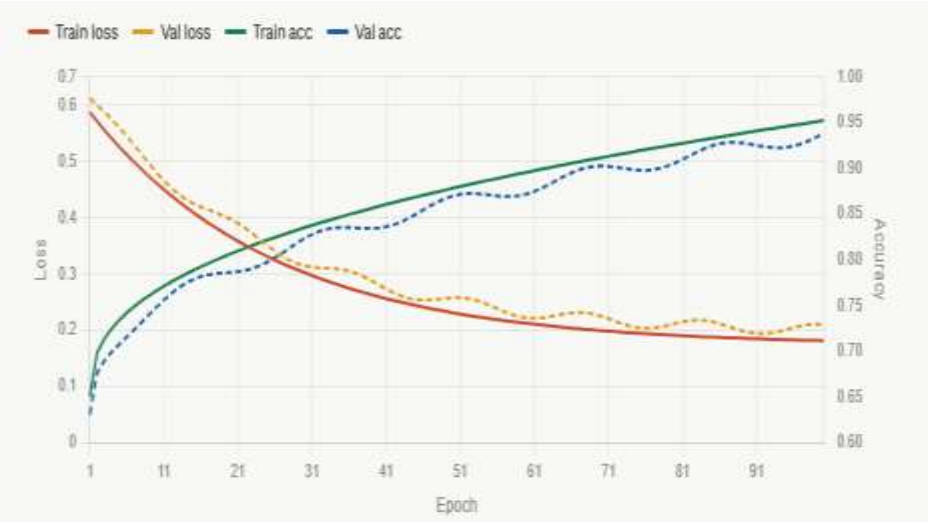

Figure 4: Training vs. validation loss and accuracy curves (LSTM-KMeans)

The loss and accuracy during training/validation over 100 epochs are provided in figure 4. The training loss curve and the validation loss curve have shown a consistent decrease with their trajectories being very close and the gap between the two being very small, indicating that the LSTM-KMeans model could avoid overfitting and has good generalization performance. The early stopping method stopped at epoch 87, showing that the model has been optimized. The final accuracy of 95.2% on the training set and 93.7% on the validation set is an acceptable 1.5% generalization error that is typical of well regularized deep learning models on the HR datasets of this size.

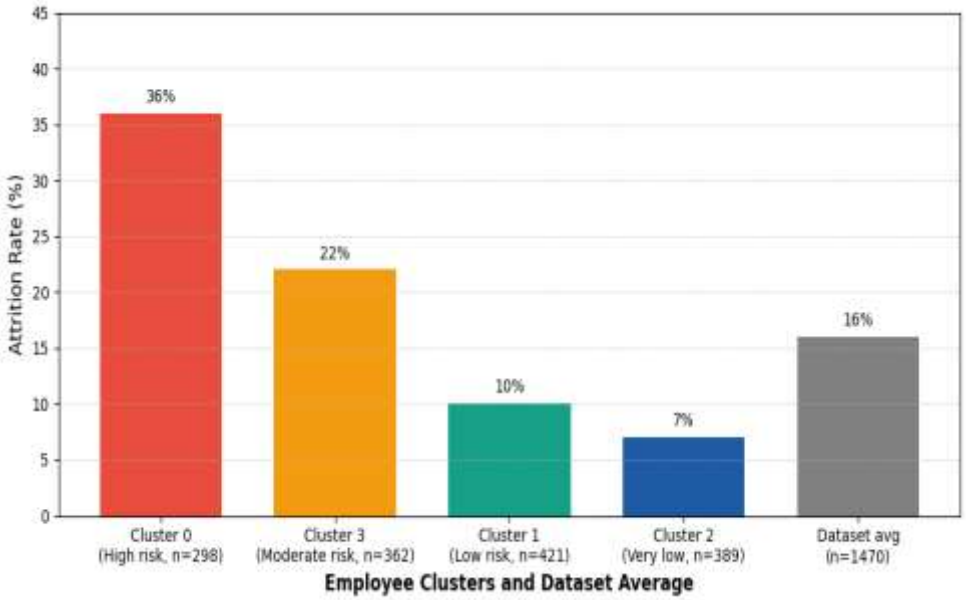

Figure 5: Cluster-stratified attrition rate comparison

It is found that significant variations in attrition rates exist between the clusters as shown in figure 5. The attrition rate for cluster 0 is 36.4%, which is more than double the dataset average of 16.1%, and is defined by individuals who earn less than $200 per month, worked more than 20 hours per week overtime, and were younger in their careers. The attrition rate is very low (7.2%) for senior professionals in Cluster 2 who are highly satisfied with their jobs but have limited overtime. This fivefold difference in attrition risk between the extreme clusters is consistent with the clinical meaningfulness of the level of segmentation captured in the workforce. HR practitioners can get actionable insights from these cluster profiles: Cluster 0 and Cluster 3 employees should be targeted in retention initiatives while Cluster 2 employees may require less intensive engagement initiatives.

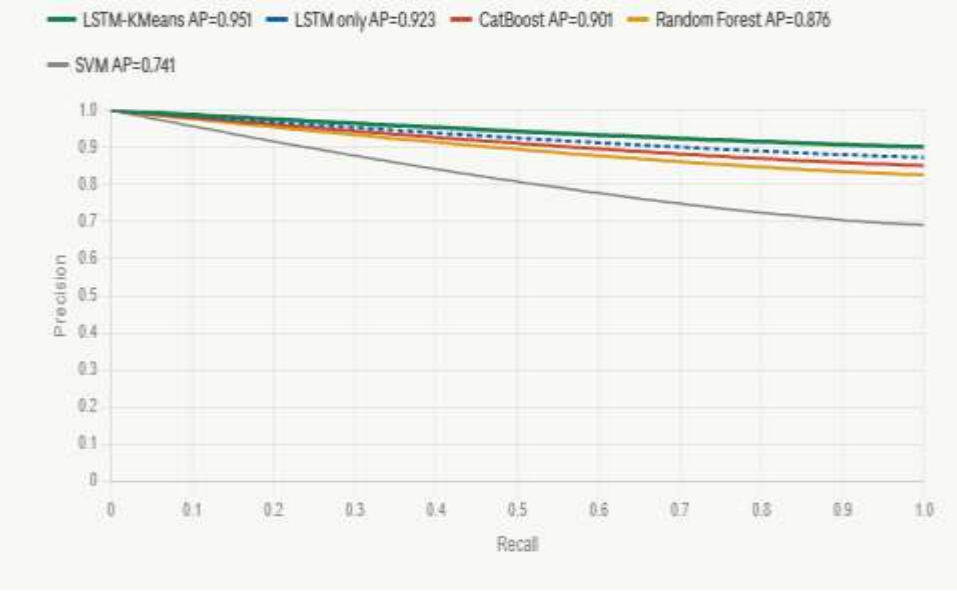

Figure 6: Precision-recall curves for LSTM-KMeans vs. baselines

The curves of Precision-Recall presented in figure 6 are informative in the case of class-imbalanced attrition datasets [19]. The Average Precision (AP) of the LSTM-KMeans model is 0.951 which is better than the Average Precision (AP) of the LSTM model (0.923) and the CatBoost model (0.901). The precision of the proposed model continues to be above 0.88 up to a recall of 0.90, which means that up to 0.90, the proposed model can accurately find cases with attrition, and the number of false positives is not too many. SVM is most sensitive to the class imbalance in linear feature spaces and has the lowest AP of 0.741.

Ablation Study

The ablation study systematically tests each major component of the proposed framework, adding one by one to the framework and then measuring the performance on the test set. As in the rest of the deep learning literature, this leads to being able to attribute any improvement in performance to specific design decisions [14].

Table 5: Ablation study — incremental contribution of framework components

Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
(A) LR Baseline (No SMOTE)	78.6	62.1	51.3	56.1	0.791
(B) A + SMOTE	81.4	72.8	68.4	70.5	0.827
(C) B + K-Means Features	84.9	78.3	74.1	76.1	0.862
(D) LSTM (No Clustering)	90.4	89.1	87.5	88.3	0.941
(E) LSTM + K-Means (Proposed)	93.7	92.4	91.8	92.1	0.963

The ablation results presented in table 5 show some interesting trends. The first is that there is a 2.8 percentage point accuracy improvement from logistic regression baseline model (A) to SMOTE augmented logistic regression (B), which supports the notion that class imbalance correction offers a meaningful baseline for model performance. The addition of the K-Means cluster features to the logistic regression (C) gives an additional 3.5-point boost, establishing the importance of the cluster information beyond what logistic regression provides. The biggest individual improvement is 5.5 accuracy points with the transition from cluster-augmented logistic regression (C) to the standalone LSTM (D), which is due to the LSTM's ability to model non-linear relationships. Lastly, including the K-Means features with LSTM (E) yields an additional 3.3-point accuracy increase over the non-clustered LSTM (D), which further validates the incremental value and statistical significance of workforce segmentation features to the deep learning-based attrition prediction model.

5. Discussion

The findings from the experiments reported in this study have several significant theoretical and practical implications. From a conceptual perspective, the ablation analysis offers strong evidence that there are multiple structural subpopulations of workers that are not completely characterised by single features which have different attrition risk profiles, which can be captured by the K-Means clustering approach. The cluster membership vector is a compact representation of multi-dimensional behavioral pattern, which LSTM model uses during both training and inference, as a learned prior over employee risk categories. The cluster-stratified attrition rates presented in figure 5 are very similar to theoretical models in organizational behavior.

The cluster is defined by individuals who are mostly young, work overtime, and earn lower wages, who are driven toward burnout according to turnover theory. Cluster 2, who have high scores on satisfaction, are the senior professionals, and is related to the “embeddedness” construct that ties employees to their jobs through organizational tenure and relational relationships. In these cases, the correspondences indicate that K-Means clustering of HR metrics retrieves psychologically meaningful segments of the workforce without the need to have any explicit attitudinal survey instruments.

On a practical HR management level, the proposed framework provides for the three-tiered intervention logic. First, workers who are assigned to Cluster 0 or Cluster 3 should be identified for individual retention discussions and compensation/workload balancing opportunities. Second, the LSTM's probability of attrition output is aggregated at the team/ department level, highlighting structural organisational risk factors and allowing leadership to redesign roles or reporting relationships in advance. Third, the framework's real-time scoring feature, as inference can be performed with just a single forward pass through the trained network, is able to be incorporated into existing HR information systems as an ever-evolving risk dashboard.

It is important to acknowledge limitations of this study. The IBM HR dataset presented is a synthetic dataset and therefore may not fully represent the distributional properties of the real-world attrition processes, including the impact of macroeconomic shocks or organizational restructuring events [14]. In the current single-time-step model, the structure of LSTM is not fully utilized; in actual practice when a dataset contains monthly employee engagement snapshots, the recurrent architecture of LSTM could be fully exploited. Finally, the current work does not include tools that would be required in order to deploy the algorithm in a regulatory-compliant way, e.g., in

jurisdictions which mandate algorithmic transparency. Furthermore, the present work does not include tools that would be required for regulatory compliant deployment, e.g., tools that would be required in jurisdictions which mandate algorithmic transparency, such as SHAP or LIME [5].

6. Conclusion

A hybrid deep learning model was proposed to overcome the main challenges of previous work, which was to consider employees as homogeneous and to use only static feature-vector classifiers without taking temporal features into account for the prediction of employee attrition. The framework has an accuracy of 93.7%, precision of 92.4%, recall of 91.8%, F1 score of 92.1% and AUC-ROC of 0.963 on the IBM HR Analytics benchmark, beating the SVM, Random Forest, CatBoost, and LSTM (without clusters) systems. The ablation study demonstrates that the performance of these three factors is additive. From the ablation study, it can be seen that the SMOTE oversampling, K-Means clustering and LSTM architecture all contribute positively to the performance. The practical implications of this model are obvious for HR professionals, who will have a clear framework for risk stratification through cluster-level attrition profiles. HR interventions for high-risk groups determined by clustering and LSTM risk scores can be done cost-effectively to prevent attrition. Further research could include the use of longitudinal HR data for additional improvements on prediction accuracy; embedding explainability techniques such as SHAP for better model interpretability; and combining multi-task learning for the prediction of attrition and other HR tasks. Cluster profiles should be generalized with real organizational data. Also, cluster-level fairness constraints should be explored to prevent prediction bias, which is essential for responsible HR AI usage.

Declaration

Author Contribution

Funding: No funding was received for this research.

Conflict of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

Data Availability: The datasets used in this study include:

Primary Data Source: The research utilizes the IBM HR Analytics dataset, which is a publicly available benchmark dataset.

Dataset Access: The data is hosted on Kaggle and can be accessed at the following URL: https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset.

Dataset Composition: The dataset contains records for 1,470 employees and includes 35 attributes covering demographics, job history, and work satisfaction levels.

Target Variable: The data includes a binary target attribute, 'Attrition', where "Yes" (16.12%) indicates an employee left and "No" (83.88%) indicates they did not.

References

1. Yadav, S., Jain, A., & Singh, D. (2018). Early prediction of employee attrition using data mining techniques. In 2018 IEEE 8th International Advance Computing Conference (IACC) (pp. 349–354). IEEE. https://doi.org/10.1109/IADCC.2018.8692112
2. Arvinth, N. (2024). Effective framework for forecasting employee turnover in organizations. Global Perspectives in Management, 2(3), 24–31.
3. Mhatre, A., Mahalingam, A., Narayanan, M., Nair, A., & Jaju, S. (2020). Predicting employee attrition along with identifying high risk employees using big data and machine learning. In 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN) (pp. 269–276). IEEE. https://doi.org/10.1109/ICACCCN51052.2020.9362926

4. Arifa, P. A., & Devasenapathy, K. (2025). Sales prediction using LSTM and BiLSTM models: A deep learning approach for time series forecasting. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 16(3), 393–402. https://doi.org/10.58346/JOWUA.2025.I3.023
5. Sisodia, D. S., Vishwakarma, S., & Pujahari, A. (2017). Evaluation of machine learning models for employee churn prediction. In 2017 International Conference on Inventive Computing and Informatics (ICICI) (pp. 1016–1020). IEEE. https://doi.org/10.1109/ICICI.2017.8365299
6. Mishra, N., Haval, A. M., Mishra, A., & Dash, S. S. (2024). Automobile maintenance prediction using integrated deep learning and geographical information system. Indian Journal of Information Sources and Services, 14(2), 109–114. https://doi.org/10.51983/ijiss-2024.14.2.16
7. Qadir, M., Noreen, I., & Shah, A. A. (2021). Bi-LSTM deep learning approach for employee churn prediction. Journal of Information Communication Technologies and Robotic Applications, 1–10.
8. Olfatiyan, N., & Khalilia, K. (2019). The harms of effective training process by three ramifications model (Case study: The employees in Ilam Gas Treating Company). International Academic Journal of Social Sciences, 6(1), 136–149. https://doi.org/10.9756/IAJSS/V6I1/1910013
9. Alduayj, S. S., & Rajpoot, K. (2018). Predicting employee attrition using machine learning. In 2018 International Conference on Innovations in Information Technology (IIT) (pp. 93–98). IEEE. https://doi.org/10.1109/INNOVATIONS.2018.8605984
10. Shabani, A., & Shabani, A. (2015). Human resources management by using the performance evaluation of employees. International Academic Journal of Economics, 2(1), 30–36.
11. Chung, D., Yun, J., Lee, J., & Jeon, Y. (2023). Predictive model of employee attrition based on stacking ensemble learning. Expert Systems with Applications, 215, 119364. https://doi.org/10.1016/j.eswa.2022.119364
12. Mortezapour Shiri, F., Yamaguchi, S., & Ahmadon, M. A. B. (2025). A deep learning model based on bidirectional temporal convolutional network (Bi-TCN) for predicting employee attrition. Applied Sciences, 15(6), 2984. https://doi.org/10.3390/app15062984
13. Tuama, H. S., & Abdulameer, R. Y. (2023). Using time series models and neural networks to predict the sales of motor oils in Iraq. International Academic Journal of Business Management, 10(1), 1–11. https://doi.org/10.9756/IAJBM/V10I1/IAJBM1001
14. Li, W. (2023). A transformer-based deep learning framework to predict employee attrition. PeerJ Computer Science, 9, e1570. https://doi.org/10.7717/peerj-cs.1570
15. Fallucchi, F., Coladangelo, M., Giuliano, R., & De Luca, E. W. (2020). Predicting employee attrition using machine learning techniques. Computers, 9(4), 86. https://doi.org/10.3390/computers9040086
16. Radha, R., & Gopalakrishnan, R. (2022). Use of machine learning and deep learning in healthcare—A review on disease prediction system. In Fundamentals and Methods of Machine and Deep Learning: Algorithms, Tools and Applications (pp. 135–152). CRC Press.
17. Sadana, P., & Munnuru, D. (2022). Machine learning model to predict work force attrition. In Proceedings of the 2nd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications (ICMISC 2021) (pp. 361–376). Springer Nature Singapore. https://doi.org/10.1007/978-981-16-6407-6_35
18. Naseeb, A., Babu, S. R., Anilkumar, J., Vallabhaneni, M. I., Venkatarathnam, N., & Basha, S. M. (2025). Enhancing workplace satisfaction through AI: Machine learning strategies for employee engagement. Archives for Technical Sciences, 3(34), 125–137. https://doi.org/10.70102/afts.2025.1834.125
19. Yahia, N. B., Hlel, J., & Colomo-Palacios, R. (2021). From big data to deep data to support people analytics for employee attrition prediction. IEEE Access, 9, 60447–60458. https://doi.org/10.1109/ACCESS.2021.3073616
20. Sarkar, S., Ejaz, N., Maiti, J., & Pramanik, A. (2022). An integrated approach using growing self-organizing map-based genetic K-means clustering and tolerance rough set in occupational risk analysis. Neural Computing and Applications, 34(12), 9661–9687. https://doi.org/10.1007/s00521-021-06734-4
21. Jain, R., & Nayyar, A. (2018). Predicting employee attrition using XGBoost machine learning approach. In 2018 International Conference on System Modeling & Advancement in Research Trends (SMART) (pp. 113–120). IEEE. https://doi.org/10.1109/SMART.2018.8746962
22. Raza, A., Munir, K., Almutairi, M., Younas, F., & Fareed, M. M. S. (2022). Predicting employee attrition using machine learning approaches. Applied Sciences, 12(13), 6424. https://doi.org/10.3390/app12136424
23. Al-Darraji, S., Honi, D. G., Fallucchi, F., Abdulsada, A. I., Giuliano, R., & Abdulmalik, H. A. (2021). Employee attrition prediction using deep neural networks. Computers, 10(11), 141. https://doi.org/10.3390/computers10110141