2024, 21(2 Special Issue): 0706-0724
https://doi.org/10.21123/bsj.2024.9788
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Simplified Novel Approach for Accurate Employee Churn Categorization using MCDM, De-Pareto Principle Approach, and Machine Learning

***Faisal Bin Al Abid*[1], *Aryati Binti Bakri*[2], *Md. Golam Rabiul Alam*[3], *Jia Uddin**[4], *Shefayatuj Johara Chowdhury*[5]**

[1]Faculty of Computing, Universiti Teknologi Malaysia, Johor Bahru, Malaysia.
[2]Faculty of Computing, Informatics Engineering, Universiti Teknologi Malaysia, Johor Bahru, Malaysia.
[3]Department of Computer Science and Engineering, Brac University, Dhaka, Bangladesh.
[4]Department of AI and Big Data, Endicott College, Woosong University, Daejeon, South Korea.
[5]Department of Computer Science and Engineering, International Islamic University Chittagong, Chittagong, Bangladesh.
*Corresponding Author.
PARS2023: Postgraduate Annual Research Seminars 2023.

Received 30/09/2023, Revised 10/02/2024, Accepted 12/02/2024, Published 25/02/2024

Abstract

Churning of employees from organizations is a serious problem. Turnover or churn of employees within an organization needs to be solved since it has negative impact on the organization. Manual detection of employee churn is quite difficult, so machine learning (ML) algorithms have been frequently used for employee churn detection as well as employee categorization according to turnover. Using Machine learning, only one study looks into the categorization of employees up to date. A novel multi-criterion decision-making approach (MCDM) coupled with DE-PARETO principle has been proposed to categorize employees. This is referred to as SNEC scheme. An AHP-TOPSIS DE-PARETO PRINCIPLE model (AHPTOPDE) has been designed that uses 2-stage MCDM scheme for categorizing employees. In 1st stage, analytic hierarchy process (AHP) has been utilized for assigning relative weights for employee accomplishment factors. In second stage, TOPSIS has been used for expressing significance of employees for performing employee categorization. A simple 20-30-50 rule in DE PARETO principle has been applied to categorize employees into three major groups namely enthusiastic, behavioral and distressed employees. Random forest algorithm is then applied as baseline algorithm to the proposed employee churn framework to predict class-wise employee churn which is tested on standard dataset of the (HRIS), the obtained results are evaluated with other ML methods. The Random Forest ML algorithm in SNEC scheme has similar or slightly better overall accuracy and MCC with significant less time complexity compared with that of ECPR scheme using CATBOOST algorithm.

Keywords: AHP-TOPSIS, DE-PARETO principle, Employee churn, MCDM, Random Forest algorithm.

Introduction

What is employee churn? What happens within an organization when employees churn? The departure of individuals and subsequently intellectual capital from an organization is known as employee churn

2024, 21(2 Special Issue): 0706-0724
https://doi.org/10.21123/bsj.2024.9788
P-ISSN: 2078-8665 - E-ISSN: 2411-7986

or turnover[1]. Employee churn can have an adverse influence on the productivity, viability, competitiveness, and benefit within an organization[2]. In this context, preventing employee churn within an organization is essential for its existence. This problem of employee turnover has become a persistent phenomenon for corporations in recent times[3]. For instance, in South East Asia, according to the Aon Hewitt poll conducted during 2015, Malaysia had voluntary employee churn rate- at 9.5% and the involuntary employee churn rate- at 6.0%, in construction industries[4]. So, solving the issue of employee churn has become a burning question now days for corporations. There are several factors that lead to employee churn, such as employment duration, years since last promotion, performance rating, working years, demographic feature[5]. These factors have influence on employee churn or turnover. Recently, since the booming of machine learning algorithms, ML has been utilized in various sectors, such as intrusion detection system for accurately predicting intrusion detection using machine learning algorithms[6]**.** This is also the scenario for employee churn prediction using various ML techniques. For example, one research aims to address employee issues concerned with rapid loss of key human resources for IT professionals using hybrid data mining techniques combined with ML clustering mechanisms[7]. Moreover, a new developed version of - weighted quadratic random forest algorithm has been used in high-dimensional unbalanced employee dataset for predicting employee churn[8]. Similarly, an efficient employee churn detection has been utilized using different ensemble methods[9]. Major factors of employee turnover are detected and developed using a characteristic model for predicting the likelihood of employee churn[10]. Up to date, only one study deals with employee categorization based on distressed, behavioral and enthusiastic employees using complex improved entropy weight method (IEWM) with TOPSIS and ML algorithms[11]. However, this approach lacks simplicity which leads to higher time complexity. SNEC scheme has been designed for employee categorization with less time complexity and slightly better or similar performance metrics compared to existing AEIM (accomplishment employee importance model).

MCDM methodologies has been exploited incrementally, particularly for Enterprise Resource Planning (ERP) software[12], health care quality, services and management[13], optimal scene for solid waste management[14]. Among various MCDM approaches, TOPSIS[15] has become a prominent method because of its high computational efficiency and straightforward mathematical structure for determining how well each criterion or alternative performs in a given scenario[16]. The efficiency of performance of alternatives is evaluated by their resemblance to the ideal or perfect solution by using TOPSIS. The best choice would be the one that was the furthest away from the worst-case scenario (NIS) and the closest to the best-case scenario (PIS). There are different ways that have been developed for finding out the relative weights of the features or dimensions in TOPSIS. Subjective weights i.e., weights given by decision makers as well as calculated objective weights based on end user's ratings are used in TOPSIS[17]. The drawback of this approach is that subjective weights given by decision makers are prone to error whereas the objective weight is calculated using complex mathematics yielding higher time complexity. Furthermore, equal weights are assigned for both the TOPSIS and modified TOPSIS methods in the simulation study[18]. When equal weights are assigned to criteria, it signifies that all criteria have the same degree of importance which may not reflect the real-world scenario. Similarly, the study in[11] uses complex mathematics in improved entropy weighted method (IEWM) to assign weights to TOPSIS for employee categorization, which definitely increases the time complexity . So, analytic hierarchy process can be used to allocate weights for criteria in TOPSIS since it is not a statistical method and can provide accurate results even with small datasets[19,20]. If the experts provide with wrong judgement, it can be found out by using consistency index (CI) or consistency ratio (CR) in AHP[21]. Moreover, AHP is considered to be a reliable method in business decision making since it deals with both quantitative and qualitative data[22].

Upon observing the employee churn dataset in Kaggle[23], it is observed that the employee churn dataset is imbalanced. Since De-pareto principle has been utilized for classifying imbalanced inventory items[24] into three classes (from most important to least important items), this concept can also be

2024, 21(2 Special Issue): 0706-0724
https://doi.org/10.21123/bsj.2024.9788
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

applied for the imbalanced Kaggle employee churn dataset with lesser time complexity.

Based on the above evidences, an MCDM coupled with machine learning based- De-Pareto principle scheme is chosen for employee churn problem. The scheme has been referred to as simplified novel employee churn (SNEC) based on TOPSIS which integrates Analytic hierarchy process (AHP) for estimating weights of criteria. In this research, firstly a novel AHPTOPSISDE is formulated for quantifying significance of employees and categorizing employees into three groups according to their accomplishment factors using De-Pareto principle. Secondly, the prediction of category wise employee churn is predicted using different tree-based classification algorithms and the results are compared. Finally, it has been shown that proposed SNEC scheme using Random Forest classification method has similar or slightly better performance metrics compared to that of ECPR scheme. Furthermore, time complexity of proposed method is significantly lower in when compared ECPR method.

Employee churn problem designed in this research addresses the following issues: 1) It provides slightly better or similar performance metrics to categorize employees based on accuracy and MCC, 2) The SNEC scheme provides less time complexity compared with existing method.

The remaining portions of research is organized as follows: related works represent review literature of employee churn problem with summarized table and existing AEIM method, working procedure of our proposed algorithm is represented in SNEC scheme, results of research last Work Assessment in terms of performance metrics and time complexity is represented in results and discussion section, and lastly conclusion section concludes the research.

Related Work

Existing studies of the researchers have focused on the employee churn prediction in human resource analysis. There have been researches focusing on the factors of voluntary employee churn based on machine learning solution[25]. 67 factors grouped in 9 factor groups were identified and among these 46 data clusters were selected for relevant decision making on employee churn. However, instead of applying machine learning methods explicitly, the research solely focuses on the theoretical aspects of employee turnover and the machine learning model.
The following discussion focus on contemporary MCDM approaches that have been directly used for forecasting employee turnover.
Ghazi, A. *et al.*[26] designed a data mining approach for recognizing the key factors of employee churn in the IT industry**.** Classification algorithms for instance random forest (RF), SVM, DT, DL, generalized linear model, etc. were used for prediction of employee churn and demographics, attitudinal characteristics, historical behavior were categorized as features for predicting employee churn. Generalized linear model had been found to have the highest accuracy compared to the other methods while detecting employee churn in this research study. However, this research study does not explicitly mention the dataset, neither deals with important performance metrics such as Matthew's correlation co-efficient.
Jain *et al.*[27] described a novel approach to categorize employees according to churn. This research used various classifications, such as RF, DT, LR, SVM, Catboost, XGBoost, etc. to categorize employees into distressed, behavioral and enthusiastic employee class. The result of the research revealed that the framework used in this research was optimum for CATBOOST algorithm. The framework used in this study had complex mathematical calculation with good accuracy but greater time complexity.
Gao *et al.*[28] proposed an efficient technique of random forest in order to predict employee churn. This study highlighted weighted quadratic random forest algorithm (WQRF) for effectively predicting high-dimensional imbalanced data. Moreover, this study focused on the significant factors of employee churn within Chinese communication enterprises. The study does not mention the range of threshold value to be applied for finding the highest weighted voting method. Moreover, this study is only applied to a specific branch of Chinese communication company.
In order to predict employee turnover among technology experts in Taiwanese companies, Fan *et*

al.[7] combined machine learning clustering analysis with heterogeneous data mining methodologies. The study dealt with employees of age range from 20-39 having the highest rate of churn to predict their churn likelihood and increase the competitiveness of organizations. Self-organization method (SOM) was used to recognize the churn trend clusters while back-propagation neural network (BPN) was utilized to find out the turnover status. However, these results were not compared with other classification methods and it was unclear whether all factors related to churn was applied for clustering and classification.
According to Canco *et al.*[29], AHP is a reliable method to explore organization's goals and can be used extensively for making decisions in business and organizations for analyzing quantitative and qualitative data. Both primary and secondary data were collected to carry out experiments. However, this study was limited to a specific context.
Brzozowski *et al.*[12] recommended MCDM methods for choosing ERP. This research identifies the significant researches in the selection of MCDM methods for ERP selection process and provides a literature analysis regarding the acceptance of various MCDM methods. No global agreed upon MCDM methods were identified to select ERP in this research.
S. Hanan *et al.*[9] proposed to develop ML models for predicting employee churn using the IBM HR Analytics. The primary aim of this research is to assist companies in improving employee satisfaction. In this study, six distinct models associated with ML, such as Decision Tree, Gradient Boosting Classifier, Logistic Regressor, Adaboost Model, and Random Forest Regressor, were used to predict employee churn. When compared to the other approaches, the logistic regression (LR) model performed the best. In order to ameliorate the machine learning model's accuracy, 3 ensemble models containing the most effective algorithms were created. The paper does not deal with datasets other than IBM HR analytic and does not consider external factors such as economic condition for detection of employee churn.

Materials and Methods

Existing AEIM Method

The existing AEIM method or accomplishment-based employee importance method deals with entropy weighted method and TOPSIS method to rank the employees and categorize employees into 3 classes using max_x and min_x interval-based scale. The flow graph of the existing categorization of the employees is shown in Fig. 1.

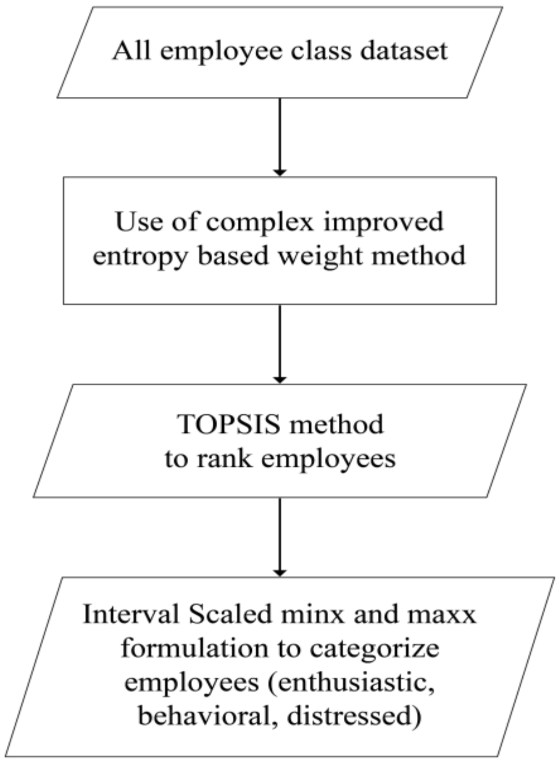

Figure 1. Flowgraph of Existing AEIM method.

Proposed SNEC Model

Fig.1 shows the whole SNEC scheme which comprises of two segments. In the 1st segment, MCDM methods such as AHP-TOPSIS is used to rank and find out the importance score of employees. In the second segment, De-pareto principle is applied to the dataset and the output of De-pareto principle is used for classifying or categorizing the employee into three classes namely distressed, behavioral and enthusiastic employees. The collection of employees Kaggle dataset is accomplished using human resource information system (HRIS) dataset which consist of 14,999 instances of the dataset with ten features. Among the 10 features included, 6 falls into the integer category, 2 belong to the floating-point type, and 2 are characterized as categorical. Detailed descriptions of these features can be found in Table 1.

Table 1. Description of employee features from Kaggle dataset

Renamed Attributes	Data type	Description
Employee Satisfaction	FL	Range 0-1.
Left	Int	Whether employee resigns or not
last Work Assessment	FL	The value range is 0-1
Project Quantity	Int	The number of projects in total for the employee
Monthly Work Hours	Int	In a month, the number of working hours is calculated as average
Employment Duration	Int	Staying duration in years
Work Accident	Int	If employee was faced with an accident or not
Promotion	Int	Had any promotion or not over 5 years
Department	Cat	Which department the employee works on
Salary	Cat	Low, medium, high: 3 levels of salary range

* FL = Floating type, Int =Integer type, Cat=Categorical type. *

Overview of AHP and TOPSIS MCDM Method

This section discusses about the concept of AHP and TOPSIS. AHP aka- analytic hierarchy process is used for ranking different criteria using consistency index as mentioned earlier. AHP is widely used for solving critical issues such as ranking the automotive vehicle comfort for seats which can be found out based on the preferences of customers[30]. The weights found from AHP can be used as an input parameter for MCDM methods like TOPSIS in order to rank the dimensions. Working flow of AHP and TOPSIS is depicted in Fig.2.

2024, 21(2 Special Issue): 0706-0724
https://doi.org/10.21123/bsj.2024.9788
P-ISSN: 2078-8665 - E-ISSN: 2411-7986

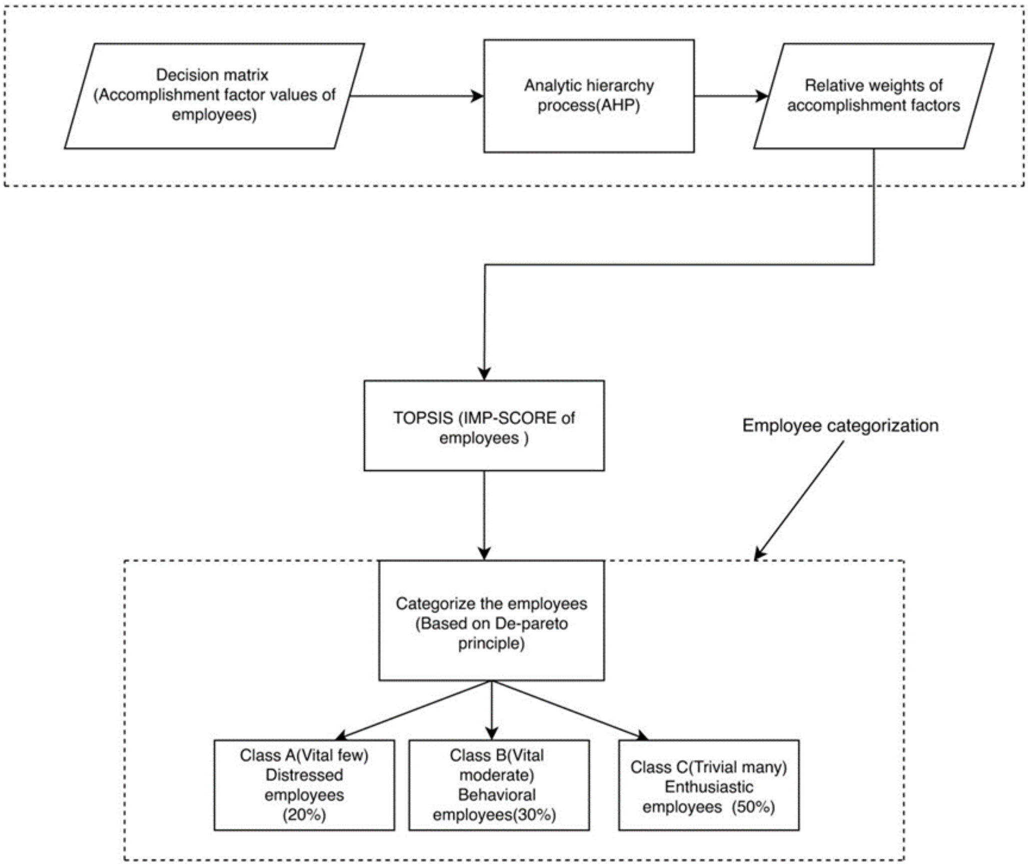

Figure 2. Conceptual work flow of the proposed SNEC scheme

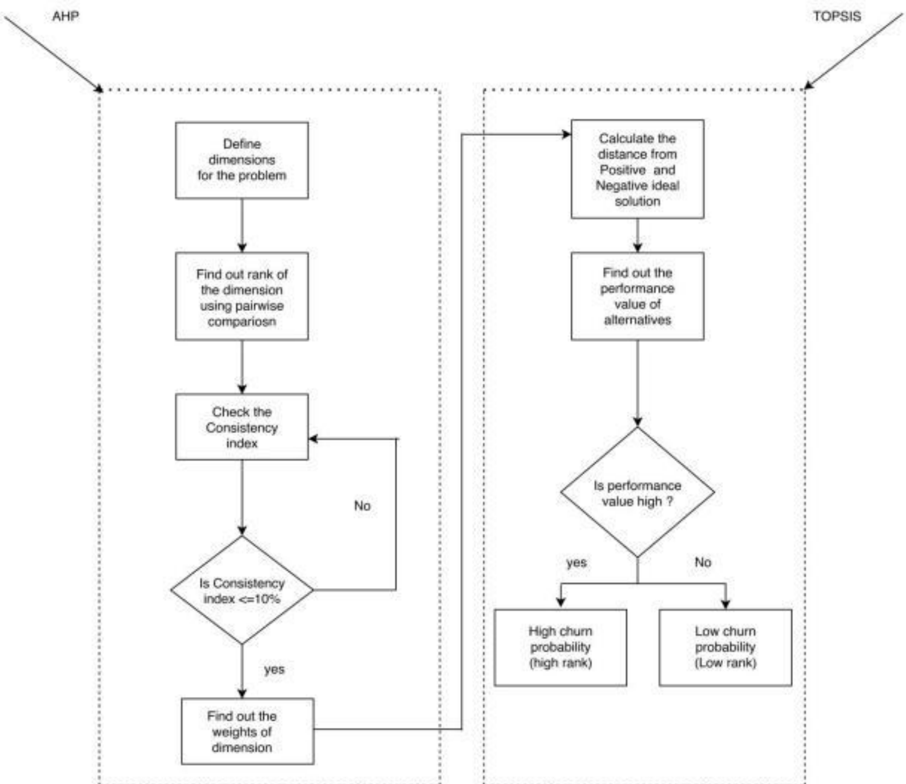

Figure 3. The working flow graph of “AHP” & “TOPSIS”

AHP-TOPSIS Method

In TOPSIS, the weights acquired from AHP technique are applied to rank employees based on their churn or turnover. Following are the detailed instructions for the AHP-TOPIS method:

Step 1: If there are K criteria, a pairwise matrix is constructed by using the following equation:

$$\boldsymbol{P_k} = \begin{bmatrix} \mathbf{p_{11}} & \mathbf{p_{12}} & \cdots & \mathbf{p_{1k}} \\ \mathbf{p_{21}} & \mathbf{p_{22}} & \cdots & \mathbf{p_{2k}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{p_{k1}} & \mathbf{p_{k2}} & \cdots & \mathbf{p_{kk}} \end{bmatrix} \qquad \text{Eq.1}$$

P is a $k \times k$ square matrix for k criteria. The experts give value from 1 to 9 based on their preferences or choices which is depicted in Table 2. Decision matrix is illustrated in Table 3 below where sample employees are considered for constructing pairwise comparison matrix.

Table 2. Preference level of experts

1	**Same preference**
3	Moderate preference
5	Strong preference
7	Very strong choice or preference
9	Extremely strongly choice or preference
2, 4, 6 or 8	Preference halfway between the integers on either side

Table 3. Pairwise comparison matrix

	Employee Satisfaction	**last Work Assessment**	**Project Quantity**	**Monthly Work Hours**	**Employment Duration**
Employee Satisfaction	1	2	2	2	5
last Work Assessment	0.5	1	2	2	5
Project Quantity	0.5	0.5	1	2	3
Monthly Work Hours	0.5	0.5	0.5	1	3
Employment Duration	0.2	0.2	0.333333	0.333333	1
Sum	2.7	4.2	5.833333	7.333333	17

Here P_{ab} denotes the kth preference of ath criterion over bth criterion. The pairwise comparison matrix given by the experts.

Step 2: The following equation calculates the geometric mean for each feature or criterion:

$$\boldsymbol{G_{ij}} = \frac{\boldsymbol{P_{ab}}}{\sum_{a=1}^{n} \boldsymbol{P_{ab}}} \quad \text{for a = 1, 2, …, k and b = 1, 2, …, k} \qquad \text{Eq.2}$$

The calculation of the geometric mean is shown in the Table 4 below for each of the criteria.

Table 4. Geometric mean of the attributes

	Employee Satisfaction	**last Work Assessment**	**Project Quantity**	**Monthly Work Hours**	**Employment Duration**
Employee Satisfaction	1 / 2.7 = 0.37037	2 / 4.2 = 0.47619	2 / 5.833333 = 0.342857	2 / 7.333333 = 0.272727	5 / 17 = 0.294118
last Work Assessment	0.5 / 2.7 = 0.185185	1 / 4.2 = 0.238095	2 / 5.833333 = 0.342857	2 / 7.333333 = 0.272727	5 / 17 = 0.294118
Project Quantity	0.5 / 2.7 = 0.185185	0.5 / 4.2 = 0.119048	1 / 5.833333 = 0.171429	2 / 7.333333 = 0.272727	3 / 17 = 0.176471
Monthly Work Hours	0.5 / 2.7 = 0.185185	0.5 / 4.2 = 0.119048	0.5 / 5.833333 = 0.085714	1 / 7.333333 = 0.136364	3 / 17 = 0.176471
Employment Duration	0.2 / 2.7 = 0.074074	0.2 / 4.2 = 0.047619	2 / 0.333333 = 0.057143	0.333333 / 7.333333 = 0.045455	1 / 17 = 0.058824
Sum	1 / 2.7 = 0.37037	2 / 4.2 = 0.47619	2 / 5.833333 = 0.342857	2 / 7.333333 = 0.272727	5 / 17 = 0.294118

The following equation determines the weight of the criteria:

$$\widehat{\boldsymbol{w_i}} = \frac{\boldsymbol{G_{i1}} + \mathbf{G_{i2}} + \cdots + \mathbf{G_{ik}}}{\boldsymbol{k}} \qquad \text{Eq.3}$$

The weight w_i is the weight of ith criteria. 5 criteria or attributes have been used here after feature extraction, the weight w_1 is calculated for job

2024, 21(2 Special Issue): 0706-0724
https://doi.org/10.21123/bsj.2024.9788
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Employee Satisfaction. Similarly, w_2 … … w_5 is calculated. The calculation is shown below:

For the criteria Employee Satisfaction,

Row sum (R1) = 0.37037+0.47619+0.342857+0.272727+0.294118 = 1.756263 Eq.4

Here, the number of criteria, n=5.

So that, w_1 = Row sum(R1)/n = 1.756263/5=0.351252581840817 Eq.5

Similarly,

Row sum (R2) = 0.185185+0.238095+0.342857+0.272727+0.294118 = 1.332982 Eq. 6

Row sum (R3) = 0.185185+0.119048+0.171429+0.272727+0.176471 = 0.92486 Eq. 7

Row sum (R4) = 0.185185+0.119048+0.085714+0.136364+0.176471 = 0.702782 Eq. 8

Row sum (R5) = 0.074074+0.047619+0.057143+0.045455+0.058824 = 0.283115 Eq. 9

Table 5. Weight of the 5 criteria after calculation for single expert

	Employee Satisfaction	last Work Assessment	Project Quantity	Monthly Work Hours	Employment Duration	Row Sum = $x_{i1} + x_{i2} + \cdots + x_{in}$	Weight = Row sum / no. of Criteria
Employee Satisfaction	0.37037	0.47619	0.342857	0.272727	0.294118	1.756263	0.351253
last Work Assessment	0.185185	0.238095	0.342857	0.272727	0.294118	1.332982	0.266596
Project Quantity	0.185185	0.119048	0.171429	0.272727	0.176471	0.924859	0.184972
Monthly Work Hours	0.185185	0.119048	0.085714	0.136364	0.176471	0.702781	0.140556
Employment Duration	0.074074	0.047619	0.057143	0.045455	0.058824	0.283114	0.056623

Similarly, weight calculation using AHP was accomplished by 4 more experts. The weight calculation using AHP for the 5 experts is given below using EX1…. EX5:

Table 6. Weight of the 5 criteria after calculation for all experts

Features	EX1	EX2	EX3	EX4	EX5
Employee Satisfaction	0.351253	0.502819	0.468392	0.513154	0.443616
last Work Assessment	0.266596	0.260231	0.268058	0.263499	0.261804
Project Quantity	0.184972	0.134350	0.143553	0.129833	0.152812
Monthly Work Hours	0.140556	0.067777	0.075858	0.061087	0.089157
Employment Duration	0.056623	0.034820	0.044138	0.032423	0.052609

Step3: Geometric mean calculation of AHP:

Geometric mean of AHP weight is computed according to the below equation:

$$\mathbf{Weight_i} = \left(\mathbf{a_{ij}}\right)^{(\mathbf{1/n})} \text{ for i, j} = 1, 2, \ldots, \text{k} \qquad \text{Eq.10}$$

For employee satisfaction, $Weight_i$ = (0.351253*0.502819*0.468392*0.513154*0.443616) ^ (1/5)=0.451834

Similarly calculate for all features.

Step4: Consistency check of experts:

Consistency ratio should be calculated to show if the response from the experts is correct or not. If the consistency rate is above 10%, the experts are given the opportunity to re-submit their answers.

The matrix P is multiplied by their respective criteria weight according to the below equation:

$\boldsymbol{P_{ij}} \times \boldsymbol{w_i}$ for i, j = 1, 2, …, k Eq.11

The weighted sum is calculated as,

$\mathbf{S_i} = \mathbf{P_{i1}}' + \mathbf{P_{i2}}' + \cdots + \mathbf{P_{ik}}'$ for i = 1,2, …, k Eq.12

Eigen vector is calculated as,

$\mathbf{Z_i} = \frac{\mathbf{S_i}}{\mathbf{w_i}}$ for i = 1,2, …, k Eq.13

Principle Eigen vector is calculated as the following formula,

$\mathbf{Z_{max}} = \frac{\mathbf{Z_1+Z_2+\cdots+Z_k}}{\mathbf{k}}$ Eq.14

The formula for consistency index is computed as:

$\textbf{Consistency Index } (\mathbf{C.I}) = \frac{\mathbf{Z_{max} - k}}{\mathbf{k - 1}}$ Eq.15

Consistency ratio is calculated as,

$\mathbf{C.R} = \frac{\textbf{Consistency Index}(\mathbf{C.I})}{\textbf{Random Index}(\mathbf{R.I})}$ Eq.16

Let us consider the criteria Employee Satisfaction and apply the formula in (11) to get:

$\mathbf{P_{11}}'$ = 1*0.351253 = 0.351253 Eq.17

$\mathbf{P_{12}}'$ = 2*0.266596 = 0.533193 Eq.18

$\mathbf{P_{13}}'$ = 2*0.184972 = 0.369944 Eq.19

$\mathbf{P_{14}}'$ = 2*0.140556 = 0.281113 Eq.20

$\mathbf{P_{15}}'$ = 5*0.056623 = 0.283114 Eq.21

The weighted sum for Employee Satisfaction is calculated as,

S_1 = 0.351253+0.533193+0.369944+0.281113+0.283114 = 1.818616 Eq.22

For the attribute Employee Satisfaction, Eigen vector Z1 is calculated as,

$\boldsymbol{Z_1} = \frac{\mathbf{S_1}}{\boldsymbol{w_1}} = \mathbf{1.818616/0.351253} = \mathbf{5.177516}$ Eq.23

In the same way, Z_2......................Z_5 is calculated using the same formula has been depicted in Table 7.

Table 7. Calculation of principle Eigen vector

	Employee Satisfaction	**last Work Assessment**	**Project Quantity**	**Monthly Work Hours**	**Employment Duration**	**Weighted Sum**	Z_i	Z_{max}
Employee Satisfaction	0.351253	0.533193	0.369944	0.281113	0.283114	1.818616	5.177516	
last Work Assessment	0.175626	0.266596	0.369944	0.281113	0.283114	1.376393	5.162833	
Project Quantity	0.175626	0.133298	0.184972	0.281113	0.169868	0.944877	5.108222	5.12233448839221
Monthly Work Hours	0.175626	0.133298	0.092486	0.140556	0.169868	0.711835	5.064414	
Employment Duration	0.070251	0.053319	0.061657	0.046852	0.056623	0.288702	5.098687	

Using Eq.14 Z_{max} has been calculated and using the equation of Eq.15 consistency index has been measured which gives away the result (C.I) = 0.0305836220980522. Now as the value of n=5, the RI value is 1.12.

So, $\text{Consistency Ratio} = \frac{\text{C.I}}{\text{R.I}} * 100\% = \frac{0.0305836220980522}{1.12} * 100\% = 0.0273068054446895 * 100\% = 2.73\%$. Since the Consistency Ratio is less than 10%, the inconsistency of the answers of the experts is acceptable.

Step 5: Enhanced modified weight of AHP:

Enhanced modified weight of AHP is calculated according to this formula:

2024, 21(2 Special Issue): 0706-0724
https://doi.org/10.21123/bsj.2024.9788
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

$$W_i = C_i * \left(\pi_{j-1}^{n} a_{ij}\right)^{(1/n)} \quad \text{Eq.24}$$

where C_i is the consistency adjustment factor of i-th criterion. The formula of adjustment factor is,

$$C_i = 1/\text{C.R} \quad \text{Eq.25}$$

Using the Eq.25, calculated the C.R is 0.045. After using Eq.24, got the final weight and normalized the weight to get final weight.

Table 8. Enhanced modified weight of AHP

Features	Enhanced modified weight	Normalized weight
Employee Satisfaction	10.04077019	0.456476202
last Work Assessment	5.867156724	0.266734261
Project Quantity	3.286760816	0.149423607
Monthly Work Hours	1.844154699	0.083839458
Employment Duration	0.957419673	0.043526471

Step 6: Let us construct a decision matrix where decision matrix D is: D (= $[d_{ij}]$) of (b×c) where b denotes number of alternatives or dimensions and c denotes number of criteria, the decision matrix is represented as follows:

$$\mathbf{D} = \begin{bmatrix} \mathbf{d_{11}} & \mathbf{d_{12}} & \cdots & \mathbf{d_{1c}} \\ \mathbf{d_{21}} & \mathbf{d_{22}} & \cdots & \mathbf{d_{2c}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{d_{b1}} & \mathbf{d_{b2}} & \cdots & \mathbf{d}_{bc} \end{bmatrix} \quad \text{Eq.26}$$

Where D_i is i^{th} selection, i = 1, ..., b; D_j is j^{th} attribute and j = 1, ..., c, and d_{ij} is D_i under criteria D_j. For instance, an example of decision matrix comprising of 10 employees has been given with the top most 5 attribute types.

Table 9. "Decision matrix" of 10 employees

Employee	last Work Assessment	Project Quantity	Monthly Work Hours	Employment Duration	Employee Satisfaction
EM1	.53	2	157	3	0.38
EM2	.86	5	262	6	0.8
EM3	.88	7	272	4	0.11
EM4	.87	5	223	5	0.72
EM5	.52	2	159	3	0.37
EM6	.5	2	153	3	0.41
EM7	.77	6	247	4	0.1
EM8	.85	5	259	5	0.92
EM9	1	5	224	5	0.89
EM10	.53	2	142	3	0.42

Step 7: Decision matrix D is normalized for standardizing the data. In order to do so, "vector normalization method" is used by constructing a uniform decision matrix U, where U(=$[U_{ij}]$) is estimated. The formula for finding out the uniform normalized value of U_{ij} is depicted below:

$$\boldsymbol{U_{ij}} = \frac{\mathbf{d}_{ij}}{\sqrt{\sum_{i=1}^{n} \mathbf{d}_{ij}^{2}}}, \boldsymbol{for\ i = 1, 2, \ldots\ldots, \mathbf{b}; j = 1, 2, \ldots\ldots \mathbf{c}} \quad \text{Eq.27}$$

For instance, let us consider the criteria or the attribute Employee Satisfaction. The attribute Employee Satisfaction is calculated for EM1 where d_{ij} = 0.38,
So, normalized decision matrix for the attribute Employee Satisfaction of

EM1= $\frac{0.38}{\sqrt{\sum_{i=1}^{n} d_{ij}^{2}}}$ = 0.38 / √ $(0.38^2+0.8^2+0.11^2+0.72^2+0.37^2+0.41^2+0.1^2+0.92^2+0.89^2+0.42^2)$ = 0.204739477.

2024, 21(2 Special Issue): 0706-0724
https://doi.org/10.21123/bsj.2024.9788
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Table 10. "Normalized decision matrix" for 10 employees.

Employee	Employee Satisfaction	last Work Assessment	Project Quantity	Monthly Work Hours	Employment Duration
EM1	0.204739477	0.222608526	0.141069123	0.230452648	0.224230528
EM2	0.431030479	0.361213835	0.352672808	0.384577031	0.448461056
EM3	0.059266691	0.369614156	0.493741931	0.399255544	0.298974037
EM4	0.387927431	0.365413996	0.352672808	0.327330832	0.373717546
EM5	0.199351596	0.218408365	0.141069123	0.233388351	0.224230528
EM6	0.22090312	0.210008043	0.141069123	0.224581243	0.224230528
EM7	0.05387881	0.323412387	0.42320737	0.362559262	0.298974037
EM8	0.495685051	0.357013674	0.352672808	0.380173477	0.373717546
EM9	0.479521408	0.420016087	0.352672808	0.328798683	0.373717546
EM10	0.226291001	0.222608526	0.141069123	0.208434879	0.224230528

Step 8: The weights obtained from AHP is calculated from step 3 and then multiplied to the components or elements of the "normalized decision matrix" where $U(=[U_{ij}])$. Weighted "normalized decision matrix", $N=([n_{ij}])$ has been computed where the normalized weighted value is calculated as the following:

$$\mathbf{n}_{ij} = \mathbf{U}_{ij} \times \boldsymbol{w}_j, \quad \mathbf{for\ i = 1, 2, \ldots, b\ and\ j = 1, 2, \ldots, c} \qquad \text{Eq.28}$$

Weights (wj) obtained from step 5 for the top most 5 criteria are given in tabular form below in table 11.

Table 11. Weights of top most criteria from step 5 of the methodology

Criteria	Weight
Employee Satisfaction	0.456476202
last Work Assessment	0.266734261
Project Quantity	0.149423607
Monthly Work Hours	0.083839458
Employment Duration	0.043526471

Using Eq.12, the value of the components of the normalized weighted value is calculated.

Table 12. Weighted normalized value for 10 employees

Employee	Employee Satisfaction	last Work Assessment	Project Quantity	Monthly Work Hours	Employment Duration
EM1	0.093458699	0.059377321	0.021079057	0.019321025	0.009759964
EM2	0.196755156	0.096348105	0.052697643	0.03224273	0.019519927
EM3	0.027053834	0.098588759	0.0737767	0.033473369	0.013013285
EM4	0.17707964	0.097468432	0.052697643	0.02744324	0.016266606
EM5	0.09099926	0.058256994	0.021079057	0.019567153	0.009759964
EM6	0.100837017	0.05601634	0.021079057	0.01882877	0.009759964
EM7	0.024594394	0.086265164	0.063237172	0.030396772	0.013013285
EM8	0.226268429	0.095227779	0.052697643	0.031873538	0.016266606
EM9	0.218890111	0.112032681	0.052697643	0.027566303	0.016266606
EM10	0.103296457	0.059377321	0.021079057	0.017475067	0.009759964

Step 9: PIS and NIS is represented as "Positive" and "Negative" ideal solution.

PIS = $\max\{n_{1j}, \ldots n_{nj}\}$ and NIS = $\min\{n_{1j}, \ldots, n_{nj}\}$, for positive accomplishment.
PIS = $\min\{n_{1j}, \ldots n_{nj}\}$ and NIS = $\max\{n_{1j}, \ldots, n_{nj}\}$, for negative accomplishment.

PIS = Positive ideal solution = employees having high turnover i.e., lesser Employee Satisfaction, last Work Assessment, Project Quantity and less Employment Duration
NIS = Negative ideal solution = employees having lower turnover i.e., greater Employee Satisfaction, last Work Assessment, Project Quantity and more Employment Duration

2024, 21(2 Special Issue): 0706-0724
https://doi.org/10.21123/bsj.2024.9788
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Table 13. PIS and NIS for employee churn

	Employee Satisfaction	last Work Assessment	Project Quantity	Monthly Work Hours	Employment Duration
PIS	0.024594394	0.05601634	0.021079057	0.033473369	0.009759964
NIS	0.226268429	0.112032681	0.0737767	0.017475067	0.019519927

Step 10: The separation measure is calculated using "Euclidean distance" in this study.

$$\mathbf{PIS(D)} = \sqrt{\sum_{i=1}^{b}\left(\mathbf{n_{ij}} - \mathbf{PIS}\right)^2} \ \ \mathbf{and} \ \ \mathbf{NIS(D)} = \sqrt{\sum_{i=1}^{b}\left(\mathbf{n_{ij}} - \mathbf{NIS}\right)^2} \qquad \text{Eq.29}$$

where j ∈ {1, 2, …, c}.

Step 11: Relative closeness co-efficient $\mathbf{CC_i}$ represents the relative closeness for ith dimension from the ideal solution (either PIS /NIS based on the scenario). Ideal solution ranges from the interval 0 to 1 and is determined by the following formula:

$$\mathbf{CC_i} = \frac{\mathbf{NIS}}{\mathbf{PIS + NIS}} \quad \mathbf{where\ i} \in \{\mathbf{1, 2, \dots \dots . . b}\} \qquad \text{Eq.30}$$

Table 14. Ranked employees based on descending order

Employee	PIS	NIS	$\mathbf{CC_i = \frac{NIS}{PIS + NIS}}$ **("Closeness Co-Efficient")**	Rank
EM7	0.052080202	0.204099901	0.796704734	1
EM3	0.067868147	0.200413263	0.747026278	2
EM5	0.067882321	0.155133001	0.695615888	3
EM1	0.070383787	0.15259986	0.684354491	4
EM6	0.077636343	0.147461861	0.655100121	5
EM10	0.080381941	0.144107569	0.641934534	6
EM4	0.161395353	0.056444039	0.259108503	7
EM2	0.179895763	0.042183524	0.189948035	8
EM8	0.207977364	0.030734857	0.12875276	9
EM9	0.204855121	0.024722123	0.107685426	10

Application of De-pareto principle for classifying employees

De-pareto principle states that Class A, which is the vital few consists of 20% of the population or item accompanying 80% of the relative frequency, Class B consists of 30% of the population accompanying 10% of the frequency and last but not the least Class C, which consists of 50% of the items or population consisting of 10% of the frequency. However, the rule of thumb states that if 20% of the features or items is above 60%, "entropy index" is less than control limit indicating that De-pareto principle can be applied[31]. Upon analyzing the data set, it can be seen that among 20% of the employees, i.e., 3000 employees (20*14999/100=3000 employees), 1839 employee has left 61.3% (1839*3000/100) has left. With this approach in principle, the De-pareto principle is applied for employee churn where closeness coefficient (CCi) is used as significance score or S-score. The alternatives or dimensions are classified in accordance of the significance score or S-score. The categorization is shown in Table 15.

Table 15. Categorization of employees according to churn

Employee Class	$\mathbf{CC_i}$ value	Employee numbers
Class A(Distressed) 20%	(0.724472-0.705616)	EM7, EM3
Class B(Behavioral) 30%	(0.695471-0.667213)	EM5, EM1, EM6
Class C(Enthusiastic) 50%	(0.657025-0.162077)	EM2, EM4, EM8, EM9, EM10

Algorithms used as Baseline

For proposed SNEC scheme, Random Forest (RF) algorithm is found to be more suited. The following factor makes Random Forest feasible for SNEC scheme. The whole procedure of the research including the SNEC approach is represented in Fig.4.

Versatility: An ensemble method is employed by Random Forest for both classification and regression purpose. Several decision trees are constructed during training of dataset. For classification, among the constructed decision trees, the highest number of outputs generated by most trees is selected to be the final outcome whereas the average prediction of trees is returned to be the final outcome for regression procedure[32]. Furthermore, it works well for different data types without the requirement for any type of scaling[33].

Coping with anomaly or noise data: RF is very good at finding outlier data such as intrusion detection system compared to other classification methods[34]. So, if there are outliers or noise in employee churn data, then random forest is the optimal choice.

Works best with imbalanced data: Since we utilized De-Pareto Principle, this concept divides the classes into three imbalanced classes targeting the most likely churned employees. Since Random Forest works best with imbalanced dataset[35], we propose Random Forest as our best choice for SNMC scheme.

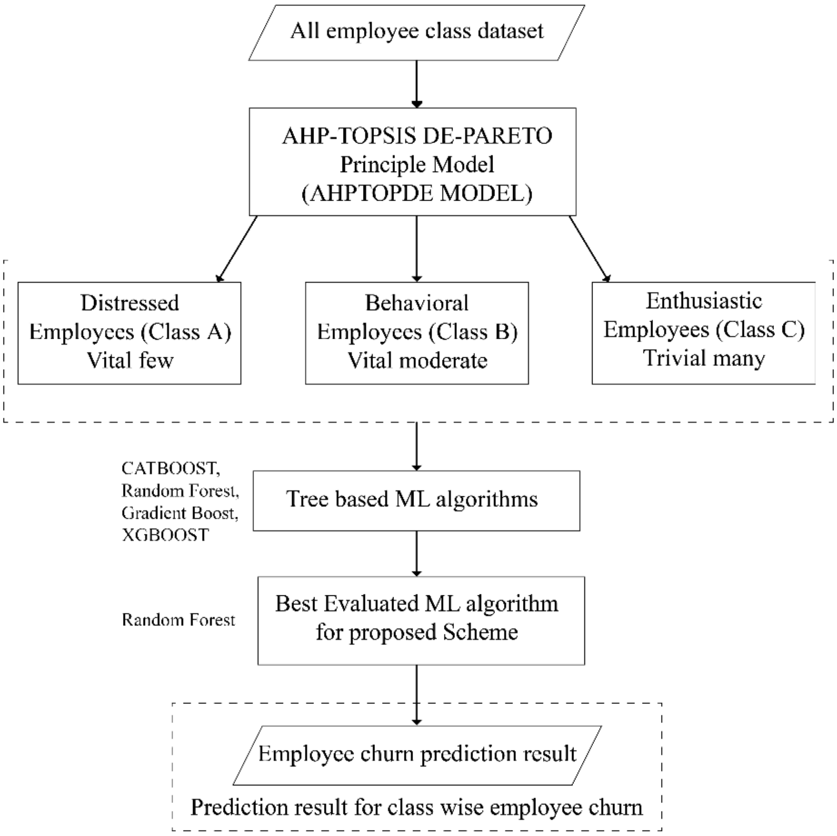

Figure 4. Conceptual flow graph for predicting churning employees

Performance Metrics

This study's primary goal is to increment proposed SNEC scheme's accuracy on the test dataset while minimizing its time complexity. The target variable left has been obtained from Kaggle dataset. The 4 terms of Confusion matrix (CM) are described herewith equation:

TRUE POSITIVE or TRPOS: TRPOS means that the algorithm detects that the employee will quit the job and in reality, the employee leaves the job.
TRUE NEGATIVE or TRNEG: TRNEG means the algorithm detects that the employee will not quit the job and in reality, the employee does not leave the job.
FALSE POSITIVE or FLPOS: FLPOS means the algorithm detects that the employee will quit the job but in reality, the employee does not leave the job.

2024, 21(2 Special Issue): 0706-0724
https://doi.org/10.21123/bsj.2024.9788
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

FALSE NEGATIVE or FLNEG: FLNEG means the algorithm detects that the employee will not quit the job but in the reality the employee quits the job.

The idea of accuracy (ACCR) for measuring overall effectiveness of classification algorithms may be represented by these four terms. The concepts are described briefly as,

$$\boldsymbol{ACCR} = \frac{\boldsymbol{TRPOS+TRNEG}}{\boldsymbol{TRPOS+TRNEG+FLPOS+FLNEG}} \quad \text{Eq.31}$$

However, in case of imbalanced dataset, accuracy may be biased. This is known as accuracy paradox[36]. In this case, Matthew's correlation co-efficient can be used as a better performance indicator. Matthew's correlation co-efficient (MCCO): The "correlation coefficient" used between predicted and original value is known as "Matthew's correlation co-efficient (MCCO)" which is mathematically formulated as,

$$\mathbf{MCCO} = \frac{TRPOS*TRNEG-FLPOS*FLNEG}{\sqrt{TRPOS+FLPOS*(TRPOS+FLNEG)*(TRNEG+FLPOS)*(TRNEG+FLNEG)}} \quad \text{Eq.32}$$

Value of MCCO ranges from -1 to +1 where -1 indicates totally wrong split prediction and +1 indicates fully accurate split prediction.

Results and Discussion

Actual and classified dataset was split into training and test dataset for experimental purposes. For convenience, the data was split into 70:30 portion for train and test dataset. The SNEC scheme had been used for obtaining prediction model by the use of train dataset and test dataset. The experimental results are depicted using the performance parameters such as Accuracy (ACCR) and time complexity. The results are shown in subsections.

Feature Extraction of the Dataset

It is noteworthy to mention that the factors used in HR dataset for employee turnover may not carry the same weight. So, feature extraction was used to find out the importance score of the features which is represented in Fig.4. Topmost 5 criteria (Employee Satisfaction, last Work Assessment, Project Quantity, Monthly Work Hours, Employment Duration) with the highest weight was selected in Fig 5. Afterwards, AHP-TOPSIS method is applied in the modified extracted features.

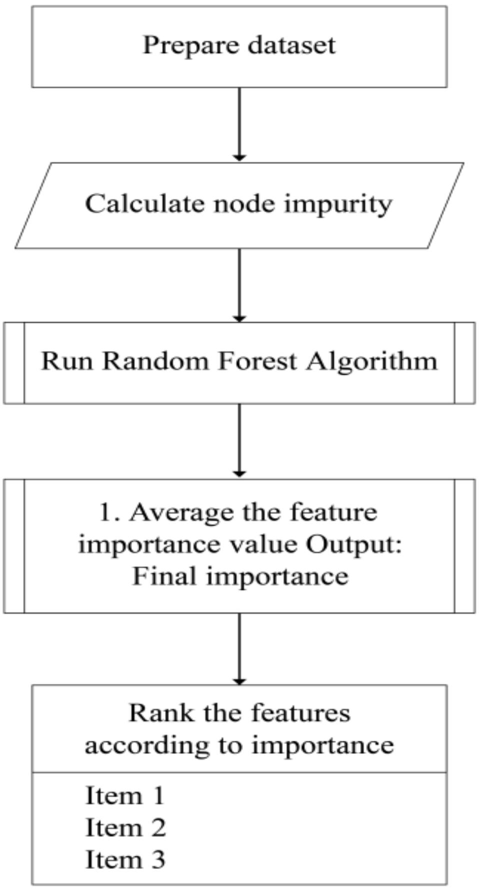

Figure 5. Feature importance algorithm

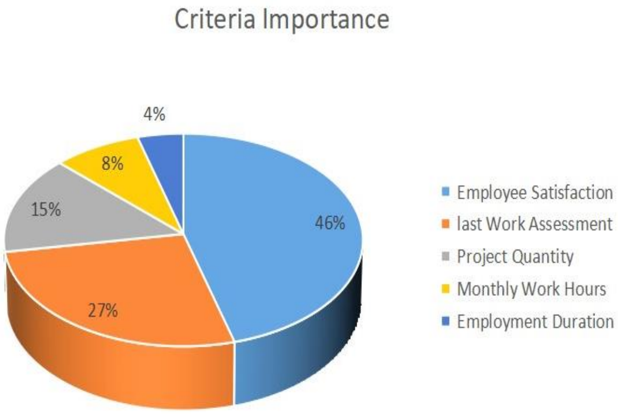

Figure 6. Top 5 criteria after feature extraction (5.1.2)

Accuracy and MCC of SNEC scheme vs AEIM Scheme

The average accuracy of different classification algorithms in SNEC approach is shown in Table 16. After applying SNEC scheme, it can be vividly visible from Table 16 that average accuracy in Random Forest is highest compared to other tree-based classifiers (99.4%). SNEC scheme works best for Random Forest because of its ability to be versatile and capability to handle imbalanced data. So, the comparison of SNEC scheme's best classifier Random Forest will be done with AEIM scheme's best classifier CATBOOST.

Table 16. Average accuracy of different classification algorithms after applying SNEC scheme

	Distress Class (Acc)	Behavioral Class (Acc)	Enthusiastic Class (Acc)	Avg Acc
RF	99.5	99.2	99.5	99.4
CATBOOST	98.5	98.8	99.9	99
XGBOOST	99	99.1	99.4	99.1
GB	98.5	98.9	98.7	98.6

Table 17. Average Accuracy (Acc) of SNEC and AEIM scheme

	Distressed class (RF_ACC)	Behavioral class (RF_ACC)	Enthusiastic class (RF_ACC)	EMclass (Avg_ACC)
SNEC (framework)	99.5	99.2	99.5	99.4
ECPR (framework)	Distressed class (CATBOOST_ACC) 98.9	Behavioral class (CATBOOST_ACC) 99.4	Enthusiastic class (CATBOOST_ACC) 98.9	EMclass (Avg_ACC) 99

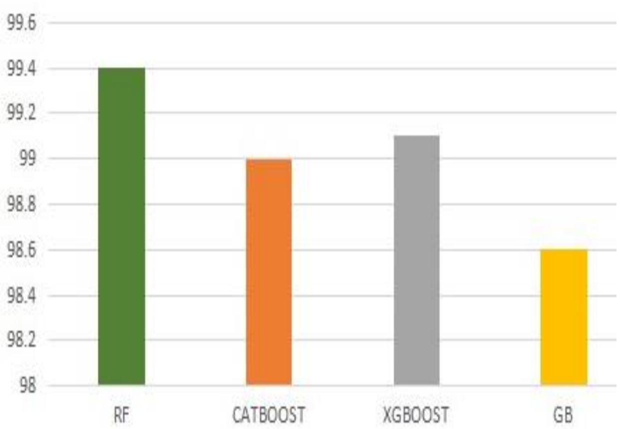

Figure 7. Average accuracy of different classification algorithms after applying SNEC scheme

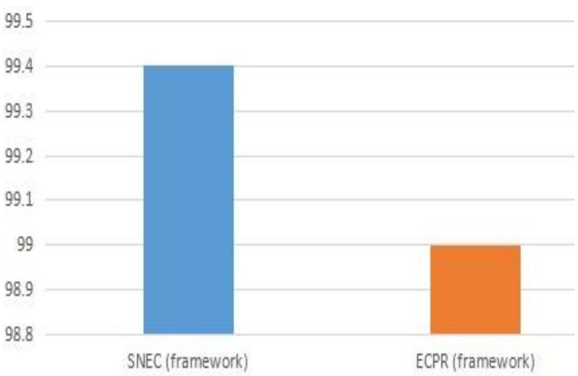

Figure 8. Average ACC of SNEC (Random Forest) and AEIM scheme (CATBOOST)

It can be clearly observed from Table 17 and Fig. 7 that SNEC scheme's Random Forest has slightly better accuracy (99.4%) compared to that of AEIM scheme's CATBOOST (99%) scheme.

2024, 21(2 Special Issue): 0706-0724
https://doi.org/10.21123/bsj.2024.9788
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Table 18. Average MCC of SNEC and AEIM scheme

	Distressed class (RF_MCC)	Behavioral class (RF_MCC)	Enthusiastic class (RF_MCC)	EMclass (Avg_MCC)
SNEC (framework)	98.9	97.4	98	98.1
ECPR (framework)	Distressed class (CATBOOST_MCC) 98	Behavioral class (CATBOOST_MCC) 98	Enthusiastic class (CATBOOST_MCC) 97	EMclass (Avg_MCC) 97.6

Similarly, MCC of SNEC scheme's best classifier Random Forest has been compared with MCC of AEIM scheme's best classifier CATBOOST. It can be clearly seen that proposed SNEC scheme (98.1%) has better MCC compared to that if existing ECPR framework (97.6%).

Time complexity of SNEC scheme and AEIM scheme

Time complexity of the two schemes is shown below in steps. It can be clearly seen that the time complexity of AEIM scheme is O(e*K^2+ KlogK) whereas the time complexity of proposed SNEC approach is O(e*K^2) where *e* is the number of employees and *K* is the accomplishment factors. This indicates that our proposed scheme has lesser time complexity compared to the AEIM scheme.

Time complexity analysis of AEIM scheme

Step1: Performance of each employee is calculated against each accomplishment factors while calculating the decision matrix. ACC_F accomplishment factors with total number of e employees has time complexity of T^1=O(ACC_F*e).
Step 2: Normalizing decision matrix by dividing each component with the sum of it's respective column, thus giving a time of T^2=O(ACC_F*e).
Step 3: Calculation of entropy weight for each ACC_F accomplishment factors, thus T^3=O(ACC_F).
Step 4: Normalize entropy weight for ACC_F accomplishment factors. T^4= O(ACC_F).

Step 5: Calculating the importance weight for each employee by calculating the weighted summation of normalized decision matrix, thus T^5=O(ACC_F*e).

Step 6: Normalization of importance weight for each employee, thus T^6 = O(ACC_F).

Step 7: Ranking the employees based on importance weight, T^7 = O(ACC_Flog ACC_F).

Step 8: All employee should be traversed for finding the maximum and minimum importance score, T^8 =O(e).

Step 9: Categorizing the employee into N classes needs each employee to be traversed to find the importance or significance score, T^9 = O(e).

Time complexity = T = $\sum_{k=1}^{K=9} T$ACC_F = O(e+e^2+ACC_F*e + ACC_F log ACC_F) = O(e*ACC_F^2+ ACC_F log ACC_F).

Time complexity analysis of SNEC scheme

Step1: AHP approach's time complexity is O(ACC_F^3+ e*ACC_F^2), where **e** = total employee and ACC_F = is accomplishment factors as mentioned earlier, T^1 = O(ACC_F^3+ e*ACC_F^2).
Step 2: TOPSIS approach has time complexity of T^2 = O(e *ACC_F^2).
Step 3: Application of De-pareto principle to classify dataset into 3 classes, T^3 = O(e*ACC_F^2).

Time complexity, T = $\sum_{j=1}^{j=3} TJ$ = O(ACC_F^3+ e*ACC_F^2)+O(e *ACC_F^2)+O(e*ACC_F ^{2}) = O(e*ACC_F ^{2})

Conclusion

This research paper focuses on proposing a simplified approach to design the SNEC framework to identify the employees who are most likely to churn. The main objective is of twofold: Firstly, design a simplified approach with less time complexity compared to entropy-based methods, Secondly, providing better performance metrics such as accuracy and MCC with the proposed approach compared to that of existing approach. It is a well-known fact that an algorithm's performance measure is computed based on the time complexity and accuracy, so proposed scheme provides an optimum trade-off between the two approaches.

Acknowledgment

The authors are thankful to the support provided by Bangabandhu Science and Technology Fellowship Trust funded by the Ministry of Science and Technology of the People's Republic of Bangladesh (Memo :39.09.0000.000.25.80.2020). The authors would also like to acknowledge the support from Universiti Teknologi Malaysia.

Authors' Declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images, that are not ours, have been included with the necessary permission for re-publication, which is attached to the manuscript.
- Ethical Clearance: Not applicable

Authors' Contribution Statement

The design, analysis and drafting of MS has been carried out by F. B. A. while the conception is accomplished by A. B. B. Acquisition of data, proof reading and interpretation has been done by G. R. A., J. U., S. J. Chowdhury, funding acquisition: J. U.

References

1. Punnoose R, Xlri -Xavier C. Prediction of Employee Turnover in Organizations using Machine Learning Algorithms. Int J Adv Res Artif Intell . 2016.https://doi.org/10.14569/IJARAI.2016.050904
2. Al-Suraihi WA, Samikon SA, Al-Suraihi AHA, Ibrahim I. Employee Turnover: Causes, Importance and Retention Strategies. Eur J Bus Manag Res. 2021 Jun 9.https://doi.org/10.24018/ejbmr.2021.6.3.893
3. Smith JD. ScholarWorks Successful Strategies for Reducing Employee Turnover in the Restaurant Industry . Walden University.
4. Ng AHH, Woo WN, Lim KY, Wong CH. Factors affecting the Staff Turnover Intention: A Case study of a Malaysian Steel Manufacturing Company. 2019.
5. Ghazi AH, Elsayed SI, Khedr AE. A Proposed Model for Predicting Employee Turnover of Information Technology Specialists Using Data Mining Techniques. Int J Electr Comput Eng Syst. 2021 Jun 21;12(2):113–21. https://doi.org/10.32985/ijeces.12.2.6
6. Mahmood RAR, Abdi AH, Hussin M. Performance Evaluation of Intrusion Detection System using Selected Features and Machine Learning Classifiers. Baghdad Sci J. 2021 Jun 20 [cited 2023 Oct 22];18(2(Suppl.)):0884–0884. https://doi.org/10.21123/bsj.2021.18.2(Suppl.).0884
7. Fan CY, Fan PS, Chan TY, Chang SH. Using hybrid data mining and machine learning clustering analysis to predict the turnover rate for technology professionals. Expert Syst Appl. 2012 Aug 1;39(10):8844–51. https://doi.org/10.1016/j.eswa.2012.02.005
8. Pratt M, Boudhane M, Cakula S. Employee Attrition Estimation Using Random Forest Algorithm. Balt J Mod Comput . 2021;9(1):49–66. https://doi.org/10.22364/bjmc.2021.9.1.04
9. Qutub A, Al-Mehmadi A, Al-Hssan M, Aljohani R, Alghamdi HS. Prediction of Employee Attrition Using Machine Learning and Ensemble Methods. Int J Mach Learn Comput. 2021 Mar;11(2):110–4. https://doi.org/10.18178/ijmlc.2021.11.2.1022
10. Zhang H, Xu L, Cheng X, Chao K, Zhao X. Analysis and Prediction of Employee Turnover Characteristics based on Machine Learning. Isc 2018 - 18th Int Symp Commun Inf Technol. 2018 Dec 24;433–7. https://doi.org/10.1109/ISCIT.2018.8587962
11. Jain N, Tomar A, Jana PK. A novel scheme for employee churn problem using multi-attribute decision making approach and machine learning. J Intell Inf Syst . 2021 Apr 1;56(2):279–302.https://doi.org/10.1007/s10844-020-00614-9
12. Brzozowski M, Birfer I. Applications of MCDM methods in the ERP system selection process in enterprises. 2017. https://www.ceeol.com/search/article-detail?id=575158
13. Glaize A, Duenas A, Di Martinelly C, Fagnot I.

Healthcare decision-making applications using multicriteria decision analysis: A scoping review. J Multi-Criteria Decis Anal . 2019 Jan 1;26(1–2):62–83. https://doi.org/10.1002/mcda.1659
14. Alhindawee Z, Jafar R, Shahin H, Awad A. Solid Waste Treatment Using Multi-Criteria Decision Support Methods Case Study Lattakia City. Baghdad Sci J . 2023 Oct 1;20(5):1575–1575. https://doi.org/10.21123/bsj.2023.7472
15. Hwang CL, Yoon K. Multiple Attribute Decision Making. 1981;186. https://doi.org/10.1007/978-3-642-48318-9
16. Yeh CH. A Problem-based Selection of Multi-attribute Decision-making Methods. Int Trans Oper Res. 2002 Mar 1;9(2):169–81. https://doi.org/10.1111/1475-3995.00348
17. Wang TC, Lee H Da. Developing a fuzzy TOPSIS approach based on subjective weights and objective weights. Expert Syst Appl. 2009 Jul 1;36(5):8980–5. https://doi.org/10.1016/j.eswa.2008.11.035
18. Chakraborty S. TOPSIS and Modified TOPSIS: A comparative analysis. Decis Anal J. 2022 Mar 1;2:100021. https://doi.org/10.1016/j.dajour.2021.100021
19. Herath G, Prato T. Using multi-criteria decision analysis in natural resource management. Using Multi-Criteria Decis Anal Nat Resour Manag. 2017 Mar 2;1–239. https://doi.org/10.4324/9781315235189
20. Duke JM, Aull-Hyde R. Identifying public preferences for land preservation using the analytic hierarchy process. Ecol Econ. 2002 Aug 1;42(1–2):131–45. https://doi.org/10.1016/S0921-8009(02)00053-8
21. Aguarón J, Escobar MT, Moreno-Jiménez JM. Reducing inconsistency measured by the geometric consistency index in the analytic hierarchy process. Eur J Oper Res. 2021 Jan 16;288(2):576–83. https://doi.org/10.1016/j.ejor.2020.06.014
22. Gupta N, Vrat P. An evaluation of alternative business excellence models using AHP. J Adv Manag Res. 2020 Mar 23;17(2):305–31. https://doi.org/10.1108/JAMR-06-2019-0101
23. HR Analytics. (n.d.). Retrieved from https://www.kaggle.com/datasets/lnvardanyan/hr-analytics#turnover.csv
24. Kartal H, Oztekin A, Gunasekaran A, Cebi F. An integrated decision analytic framework of machine learning with multi-criteria decision making for multi-attribute inventory classification. Comput Ind Eng. 2016 Nov 1;101:599–613. https://doi.org/10.1016/j.cie.2016.06.004
25. Judrups J, Cinks R, Birzniece I, Andersone I. Machine learning based solution for predicting voluntary employee turnover in organization. In 2021. https://doi.org/10.22616/ERDev.2021.20.TF296
26. Ghazi AH, Elsayed SI, Khedr AE. A Proposed Model for Predicting Employee Turnover of Information Technology Specialists Using Data Mining Techniques. Int J Electr Comput Eng Syst. 2021 Jun 21;12(2):113–21. https://doi.org/10.32985/ijeces.12.2.6
27. Jain N, Tomar A, Jana PK. A novel scheme for employee churn problem using multi-attribute decision making approach and machine learning. J Intell Inf Syst. 2021 Apr 1;56(2):279–302. https://doi.org/10.1007/s10844-020-00614-9
28. Gao X, Wen J, Zhang C. An Improved Random Forest Algorithm for Predicting Employee Turnover. Math Probl Eng. 2019;2019.https://doi.org/10.1155/2019/4140707
29. Canco I, Kruja D, Iancu T. AHP, a Reliable Method for Quality Decision Making: A Case Study in Business. Sustain 2021, Vol 13, Page 13932. 2021 Dec 16;13(24):13932. https://doi.org/10.3390/su132413932
30. Fazlollahtabar H. A subjective framework for seat comfort based on a heuristic multi criteria decision making technique and anthropometry. Appl Ergon. 2010 Dec 1;42(1):16–28.
31. Grosfeld-Nir A, Ronen B, Kozlovsky N. The Pareto managerial principle: when does it apply? Int J Prod Res. 2007 May;45(10):2317–25. https://doi.org/10.1080/00207540600818203
32. Ho TK. The random subspace method for constructing decision forests. IEEE Trans Pattern Anal Mach Intell. 1998;20(8):832–44. https://doi.org/10.1109/34.709601
33. Computer Science Journals | IJACSA | Scopus Indexed Journal. https://thesai.org/Publications/IJACSA
34. Biswas P, Samanta T. Anomaly detection using ensemble random forest in wireless sensor network. Int J Inf Technol. 2021 Oct 1;13(5):2043–52. https://doi.org/10.1007/s41870-021-00717-8
35. Chao Chen BL. Using Random Forest to Learn Imbalanced Data. 2004.
36. Accuracy Paradox. “If you don’t know anything about… | by Tejumade Afonja | Towards Data Science. https://towardsdatascience.com/accuracy-paradox-897a69e2dd9b

2024, 21(2 Special Issue): 0706-0724
https://doi.org/10.21123/bsj.2024.9788
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

نهج مبتكر مبسط للتصنيف الدقيق لحالات تخلف الموظفين عن العمل باستخدامMCDM ، ونهج مبدأDe-Pareto ، والتعلم الآلي

فيصل بن العابد[1]، أرياتي بنتي بكري[2]، دكتور غلام ربيع العلم[3]، جيا الدين[4]، شفاياتوج جوهرة شودري[5]

[1]كلية الحاسبات، الجامعة التكنولوجية الماليزية، جوهور باهرو، ماليزيا.
[2]كلية الحاسبات، الهندسة المعلوماتية، الجامعة التكنولوجية الماليزية، جوهور باهرو، ماليزيا.
[3]قسم علوم وهندسة الحاسوب، جامعة براك، دكا، بنغلاديش.
[4]قسم الذكاء الاصطناعي والبيانات الضخمة، كلية إنديكوت، جامعة ووسونغ، دايجون، كوريا الجنوبية.
[5]قسم علوم وهندسة الحاسوب، الجامعة الإسلامية العالمية شيتاغونغ، شيتاغونغ، بنغلاديش.

الخلاصة

يعد طرد الموظفين من المنظمات مشكلة خطيرة. يجب حل مشكلة تدوير الموظفين أو اطردهم من داخل المنظمة نظرًا لأن لها تأثيرًا سلبيًا على المنظمة. يعد الاكتشاف اليدوي لتسرب الموظفين أمرًا صعبًا للغاية، لذلك تم استخدام خوارزميات التعلم الآلي ML بشكل متكرر لاكتشاف تسرب الموظفين بالإضافة إلى تصنيف الموظفين وفقًا للاستبدالهم.
باستخدام التعلم الآلي، بحثت دراسة واحدة فقط في تصنيف الموظفين حتى الآن. تم اقتراح نهج جديد لاتخاذ القرار متعدد المعايير MCDM إلى جانب مبدأ DE-PARETO لتصنيف الموظفين. يشار إلى هذا باسم مخطط SNEC .
تم تصميم نموذج AHP-TOPSIS DE-PARETO PRINCIPLE (AHPTOPDE) الذي يستخدم نظام MCDM على مرحلتين لتصنيف الموظفين. في المرحلة الأولى، تم استخدام عملية التسلسل الهرمي التحليلي AHP لتعيين الأوزان النسبية لعوامل إنجاز الموظف. في المرحلة الثانية، تم استخدام TOPSIS للتعبير عن أهمية الموظفين في إجراء تصنيف الموظفين. تم تطبيق قاعدة 20-30-50 البسيطة في مبدأ DE PARETO لتصنيف الموظفين إلى ثلاث مجموعات رئيسية وهي الموظفون المتحمسون والسلوكيون والمضطربون. يتم بعد ذلك تطبيق خوارزمية الغابة العشوائية كخوارزمية أساسية لإطار عمل الموظفين المقترح للتنبؤ بخسارة الموظفين على أساس الفصل والذي يتم اختباره على مجموعة بيانات قياسية لنظام معلومات الموارد البشرية HRIS، ويتم تقييم النتائج التي تم الحصول عليها باستخدام طرق تعلم الآلة الأخرى. تتمتع خوارزمية Random Forest ML في مخطط SNEC بدقة إجمالية مماثلة أو أفضل قليلاً وMCC مع تعقيد زمني أقل مقارنةً بمخطط ECPR باستخدام خوارزمية CATBOOST

الكلمات المفتاحية: AHP-TOPSIS، مبدأ DE-PARETO، تقلب الموظفين، MCDM، خوارزمية الغابة العشوائية.