DOI: 10.21917/ijdsml.2026.0191

DISCOVER EMPLOYEE SUSTAINABILITY UNDERSTANDINGS THROUGH TEXT-DRIVEN HR ANALYTICS

Nasreen Nasar[1] and Saiyed Umer[2]
[1]*Department of Management and Business Administration, Aliah University, India*
[2]*Department of Computer Science and Engineering, Aliah University, India*

Abstract
The development of electronic Human Resource Management (e-HRM) has provided an opportunity to make decisions on the basis of data, particularly with the assistance of sentiment analysis. The paper presents a full system of examining the emotions and attitudes of the employees towards the company with a sustainability angle. The system combines employee text and number based-data to provide an exhaustive illustration of the behavior of the employees at the workplace. The data is cleaned, important features are extracted and finally models are trained along two distinct paths: one with numbers and one with text. These models are subsequently combined to form one predictive model which produces better results through the use of both kinds of data. Two standard datasets are used to test the system, and the findings demonstrate that the combination of text-based and numerical analysis of HR can be used to identify employee engagement, satisfaction and the possibility of them leaving the company. The approach can assist in developing superior HR practices to facilitate long-term development and intelligent, emotional management of human resources.

Keywords:
e-HRM, HR Analytics, Text-based, Employee Sustainability

1. INTRODUCTION

Human Resource Management System (HRMS) is significant in learning about the ability of the sustainable employees based on the use of data to enhance the degree of well, perform and remain involved in the long term [1]. It possesses such tools as performance tracking, provision of learning opportunities and monitoring employee wellness. These features assist businesses to discover how content their employees are, how they work and balance work and personal life, and their career development. This will assist in creating a man workforce that would continue to develop and remain with the firm in the long-term [2]. HRMS also uses data on events such as employee attrition, absenteeism, and skill utilization to develop practical talent retention strategies and workforce empowerment. This relates the way people develop to the larger ambitions of the company in the direction of sustainability [3]. As an illustration, the HRMS systems monitor the level of engagement of employees, the rate of their job advancement, and the degree of their participation in training. This will allow the HR leaders to identify where the workers may require assistance to continue performing well and remain healthy. Such close monitoring also assists the employees to be familiar with the changes in the company and to build a culture of never ending learning and ever growing talent [4].

Moreover, the HRMS is useful to make the management of the workforce more unbiased and transparent, which is significant in maintaining good long-term relations with the employees [5]. Reducing biases in promotions and rewards can make the workplace more inclusive and equal through the use of such tools as skill-matching modules and automated performance reviews. These equitable practices have direct influence on trust in the organization and satisfaction of employees that are very important to human capital long-term sustainability. Moreover, advanced technologies, such as artificial intelligence (AI) and the Internet of Things (IoT) can be used by the HRMS to offer new opportunities to learn more about how the workplace functions, the way employees act, and what they embrace using advanced technologies. As an example, IoT can be used to gather real-time information about such aspects as workplace comfort and usage, and AI may provide individual career development ideas. According to [5], HRMS is a major aspect of sustainable workforce systems as these technology integrations will establish an environment that fosters personal development and organizational success in the long run.

HR relies on text-based analytics to identify the hidden evidence of how employees perceive their work and the level of engagement, as well as how sustainable their performance is. This approach examines unstructured texts emerging in such locations as internal messages, performance appraisals, exit interviews, and employee surveys. It applies methods such as sentiment analysis, topic modeling, and aspect-based sentiment analysis (ABSA) in trying to understand the data in a better way. ABSA can help the HR professionals analyze the responses to open-ended questions in the survey and locate valuable issues such as workload, recognition, and the way people get along with one another. They are also able to read the feeling behind such topics hence giving them an insight on what is making workers stressed or happy. The insights come in handy by assisting companies to ensure that employees are retained longer and perform well. This may involve such items as special training, managerial support or mental health resources. Recent surveys such as the one conducted [6] demonstrate that ABSA is effective in obtaining detailed and significant information regarding employee surveys, which will assist with the various phases of an employee career management. Also, real-time sentiment tracking throughout the internal communications enhances the prompt detection of burnout risks and employee morale shifts, which enables the HR strategies to align with the changing needs of employees and the sustainability goal of the whole organization. According to these results, the input of this paper can be summarized as follows:

- An HR analytics framework has been developed that integrates both textual and non-textual data to uncover insights into employee sustainability.
- The proposed approach effectively derives and analyzes employee reviews to assess perceptions of their organizations.
- An ensemble of efficient techniques has been employed to represent the textual content of employee attitudes, enhancing the quality of analysis.

- The resulting prediction models have been validated on multiple real-world employee review datasets collected from different organizations.

The organization of this work is as follows: Section 2 discusses the dataset curations from different sources of disclosure. The methodology implemented for the proposed work has been explained and represented in Section 3. The results and discussion has been demonstrated in Section 4. The findings of this work have been concluded in Section 5.

2. DATA COLLECTION

- **Dataset-1:** This dataset is collected from Kaggle, which contains 839555 samples of employee reviews, where the features are text and Emotion-label. In this dataset, the class imbalance results from the wide range of emotions that people experience, which are hard to adequately capture in data. The purpose of this Emotion Classification Dataset is to facilitate emotion analysis research. It contains a diverse range of texts that are labelled with the feelings they express, such as happy,hate, and sad. Developing an effective model to identify emotions in text is the aim. The characteristics of these features are shown in Table.1.

Table.1. Some samples from Dataset-1

Text	Emotion
i seriously hate one subject to death but now i feel reluctant to drop it	hate
im so full of life i feel appalled	neutral
i sit here to write i start to dig out my feelings and i think that i am afraid to accept the possibility that he might not make it	neutral
ive been really angry with r and i feel like an idiot for trusting him in the first place	anger
i feel suspicious if there is no one outside like the rapture has happened or something	neutral
i feel jealous because i wanted that kind of love the true connection between two souls and i wanted that	love

(b) Dataset-2: This dataset is collected from Kaggle, 30336 samples of this dataset have been considered for this work. Each sample has 17 features which are [‘ID’, ‘Place’, ‘location’, ‘date’, ‘status’, ‘job_title’, ‘summary’, ‘positives’, ‘negatives’, ‘advice_to_mgmt’, ‘score_1’, ‘score_2’, ‘score_3’, ‘score_4’, ‘score_5’, ‘score_6’, ‘overall’]. The characteristics of these features are shown in Table.2.

Table.2: Some samples from Dataset-2

Column Name	Data Type	Description (example meaning)
ID	int64	Unique identifier for each record
Place	object	Company / branch / location name
location	object	Geographical location (city, country)
date	object	Date of review (string format)
status	object	Reviewer’s employment status
		(“Current Employee”, “Former Employee”)
job_title	object	Job designation of the reviewer
summary	object	Short summary / headline of the review
positives	object	Positive aspects mentioned
negatives	object	Negative aspects mentioned
advice_to_mgmt	object	Suggestions for management
score_1	float64	Rating score (e.g., work-life balance)
score_2	float64	Rating score (e.g., compensation/benefits)
score_3	float64	Rating score (e.g., job security)
score_4	float64	Rating score (e.g., management)
score_5	float64	Rating score (e.g., culture/values)
score_6	int64	Another rating (likely categorical, e.g., promotion opportunities)
overall	float64	Overall rating (average or weighted score)

3. PROPOSED METHODOLOGY

Organizations can identify and address sustainability initiatives in employee communications by utilizing text-driven analytics, such as sentiment analysis and contextual sentence classification, which aligns HR procedures with more general environmental, social, and governance (ESG) objectives [7]. E-HRM (Electronic Human Resource Management) systems, as discussed in [8], help organizations become more sustainable by using advanced analytics and AI tools to find hidden insights from employee data. AI-powered HR solutions use machine learning, natural language processing, and predictive analytics to make routine HR tasks easier, improve how companies manage talent, and help leaders make better decisions. These systems create a lot of digital data from activities like e-performance, e-learning, e-compensation, e-selection, and e-recruitment. These activities involve things like feedback, performance records, training sessions, salary preferences, and messages from job applicants. Companies can use text analytics and natural language processing to analyze employee sentiment from this data, as noted by [9]. This helps understand employees’ attitudes, satisfaction, and concerns. Managers can use e-performance tools to track peer reviews and personal evaluations, and through sentiment analysis, they can spot signs of motivation or dissatisfaction. HR is then able to offer or modify resources along the learning challenges and level of engagement of the employees as in online forums and training exercises. The evaluation of compensation can indicate that the staff members are satisfied with the benefits or they feel that they are not unfairly treated, which may assist in adjusting the pay policy. It is also possible to enhance employer branding and hiring processes through the analysis of job applicant feedback and communications to demonstrate the perception of the company. Through such E-HRM activities, the e-sentiment analysis can provide real time information about the risks of losing employees, their engagement and general contentment. This will assist in developing long term HR plans that are in tandem with the objectives of the firm. The Fig.1 depicts the e-HRM activities that are associated with e-sentiment.

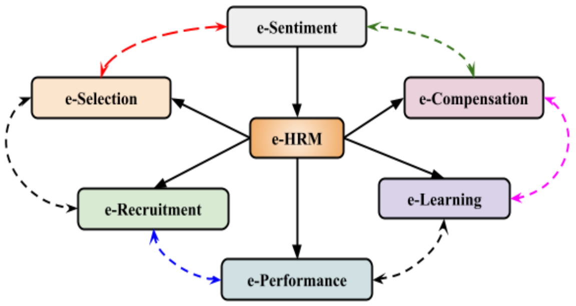


Fig.1. e-HRM activities for the proposed workforce engagement

Sentiment analysis in e-recruitment and e-selection assists in enhancing employer branding and streamlines the hiring process. It also improves the HR decision making using data. The given e-sentiment analysis approach to the e-HRM system advances this concept. It does not simply assess job seekers to determine the information about employee sustainability by analyzing HR textually. This relationship assists the organizations to know employee participation, contentment, and turnover danger in real-time, parallel workforce long-term productivity with company long-term aims and enhance strategic human resources planning. In this work, e-sentiment analysis of the e-HRM system is carried out on the basis of HR data in numerical and textual forms. These types of data are not similar and hence they are treated differently. Numerical data are prepared and standardized, and textual data are prepared into numerical representation which a computer can comprehend using text processing and embedding techniques [10]. All data forms are then trained to give different models, which are referred to as the number-model and the text-model. In order to determine the analysis more valid, the system integrates the findings of both models where either of them is not powerful enough. Moreover, when textual data is the only one which is available, the system may also carry useful sentiment information. This strategy is versatile and enjoys the ability to combine both kinds of data enabling the system to efficiently analyze employee input and aid in decision-making based on data to manage the workforce and attain organizational sustainability. The flow diagram of the proposed method is shown in Fig.2.

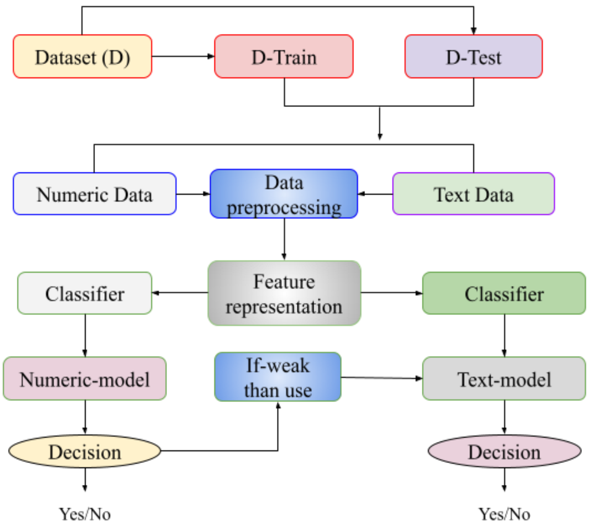


Fig.2. Block diagram of the proposed system

The process of preparing numeric data starts by dealing with missing or null values. Then, the features are split into the ones used for learning and the target label. To build the predictive models, several classifiers are used, such as Support Vector Classifier with a radial basis function kernel, Random Forest Classifier, AdaBoost Classifier, K-Nearest Neighbors Classifier, Decision Tree Classifier, Gaussian Naïve Bayes, and Quadratic Discriminant Analysis. Each model is tested on sample data to check how well it performs. The model that works the best is picked and used for the final numeric model.

- **Text preprocessing**: In the initial phase of text preprocessing, the corpus consisted of employee review comments detailing their workplace experiences, attitudes, suggestions, and remarks concerning their current or former organizations. To prepare this unstructured text data for analysis, a series of standard preprocessing steps are employed [11]. The procedure included the removal of special characters and punctuation, tokenization, the conversion of all text to lowercase, the elimination of stopwords, and feature reduction via lemmatization. The steps for these text preprocessing tasks have been demonstrated with an example in Table.3.

Table.3. Text preprocessing steps required for this work

Step	Processing	Outcome
Input: “The majestic, centuries-old oak tree, which had withstood countless storms, droughts, and even a lightning strike, stood proudly at the crest of the wind-swept hill, its gnarled branches reaching out like weary arms against the fading twilight sky.”		
1	Remove Special Characters: This removes non-alphanumeric characters.	The majestic centuries old oak tree which had withstood countless storms droughts and even a lightning strike stood proudly at the crest of the wind swept hill its gnarled branches reaching out like weary arms against the fading twilight sky
2	Perform Tokenization: This splits the text into individual words (tokens).	[‘The’, ‘majestic’, ‘centuries’, ‘old’, ‘oak’, ‘tree’, ‘which’, ‘had’, ‘withstood’, ‘countless’, ‘storms’, ‘droughts’, ‘and’, ‘even’, ‘a’, ‘lightning’, ‘strike’, ‘stood’, ‘proudly’, ‘at’, ‘the’, ‘crest’, ‘of’, ‘the’, ‘wind’, ‘swept’, ‘hill’, ‘its’, ‘gnarled’, ‘branches’, ‘reaching’, ‘out’, ‘like’, ‘weary’, ‘arms’, ‘against’, ‘the’, ‘fading’, ‘twilight’, ‘sky’]
3	Converting all Characters to Lowercase: This converts all uppercase letters to lowercase.	[‘the’, ‘majestic’, ‘centuries’, ‘old’, ‘oak’, ‘tree’, ‘which’, ‘had’, ‘withstood’, ‘countless’, ‘storms’, ‘droughts’, ‘and’, ‘even’, ‘a’, ‘lightning’, ‘strike’, ‘stood’, ‘proudly’, ‘at’, ‘the’, ‘crest’, ‘of’, ‘the’, ‘wind’, ‘swept’, ‘hill’, ‘its’, ‘gnarled’, ‘branches’, ‘reaching’, ‘out’, ‘like’, ‘weary’, ‘arms’, ‘against’, ‘the’, ‘fading’, ‘twilight’, ‘sky’]

4	Removing Punctuations: This ensures any remaining punctuation is removed.	[‘the’, ‘majestic’, ‘centuries’, ‘old’, ‘oak’, ‘tree’, ‘which’, ‘had’, ‘withstood’, ‘countless’, ‘storms’, ‘droughts’, ‘and’, ‘even’, ‘a’, ‘lightning’, ‘strike’, ‘stood’, ‘proudly’, ‘at’, ‘the’, ‘crest’, ‘of’, ‘the’, ‘wind’, ‘swept’, ‘hill’, ‘its’, ‘gnarled’, ‘branches’, ‘reaching’, ‘out’, ‘like’, ‘weary’, ‘arms’, ‘against’, ‘the’, ‘fading’, ‘twilight’, ‘sky’]
5	Removing Stopwords: Removes common words (e.g., “the”, “and”, “at”, “of”).	[‘majestic’, ‘centuries’, ‘old’, ‘oak’, ‘tree’, ‘withstood’, ‘countless’, ‘storms’, ‘droughts’, ‘even’, ‘lightning’, ‘strike’, ‘stood’, ‘proudly’, ‘crest’, ‘wind’, ‘swept’, ‘hill’, ‘gnarled’, ‘branches’, ‘reaching’, ‘like’, ‘weary’, ‘arms’, ‘against’, ‘fading’, ‘twilight’, ‘sky’]
6	Perform Stemming vs. Lemmatization: It chops word endings to reach a root form.	[‘majest’, ‘centuri’, ‘old’, ‘oak’, ‘tree’, ‘withstood’, ‘countless’, ‘storm’, ‘drought’, ‘even’, ‘lightn’, ‘strike’, ‘stood’, ‘proudli’, ‘crest’, ‘wind’, ‘swept’, ‘hill’, ‘gnarl’, ‘branch’, ‘reach’, ‘like’, ‘weari’, ‘arm’, ‘against’, ‘fade’, ‘twilight’, ‘sky’]
7	Lemmatization (using WordNet Lemmatizer): It uses a vocabulary and morphological analysis to return the base or dictionary form of a word (the lemma), which is a real word.	[‘majestic’, ‘century’, ‘old’, ‘oak’, ‘tree’, ‘withstand’, ‘countless’, ‘storm’, ‘drought’, ‘even’, ‘lightning’, ‘strike’, ‘stand’, ‘proudly’, ‘crest’, ‘wind’, ‘swept’, ‘hill’, ‘gnarled’, ‘branch’, ‘reach’, ‘like’, ‘weary’, ‘arm’, ‘against’, ‘fading’, ‘twilight’, ‘sky’]

- **Feature extraction and classification**: In order to convert unstructured textual data into a numerical format appropriate for machine learning models and computational analysis, text feature extraction is a crucial step. It entails recognizing and expressing the main linguistic, semantic, and syntactic characteristics of a text that best convey its context and underlying meaning. Common methods include word embeddings such as Word2Vec, GloVe, or FastText where words are represented as dense vectors depending on similarity at context, as well as traditional methods such as Bag of Words (BoW) and Term Frequency–Inverse Document Frequency (TF-IDF). Deep contextual knowledge is used in order to process deep contextual understanding to extract rich semantic features using models that are more sophisticated such as BERT and other transformer-based embeddings. With the ability to make algorithms more precise and effective at reading the patterns in texts, feature extraction allows enhancing the performance of tasks related to natural language processing, including text classification, sentiment analysis, and topic modelling. The Fig.3 has shown the working principle of the text-based feature extraction.

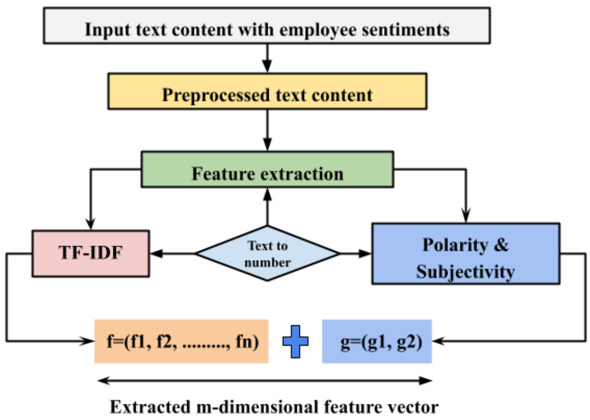


Fig.3. The working flow diagram of feature extraction employed in this study

- **TF-IDF method:** Here, the Text is converted into numerical feature vectors by the TF-IDF (Term Frequency–Inverse Document Frequency) Vectorizer, which indicates a word’s significance to a document in a collection. The two primary elements of its operation are Inverse Document Frequency (IDF) and Term Frequency (TF). The frequency with which a term p occurs in a specific document q is measured by TF [12]. It measures a word’s local significance in that document and is provided by Total terms in document q divided by the number of times term p appears in document q is equal to $TF(p,q)$. IDF calculates the term p’s importance throughout the entire set of documents. Words that appear frequently in documents have lower weights because they are less informative. It is described as: IDF(p)=log(D/(1+d)), where D be the total number of documents, and d be the number of documents that contain p. Hence, TF-IDF is given by TF-IDF(p,q)=TF(p,q)xIDF(p).
- **Polarity:** A text’s sentiment orientation, whether positive, negative, or neutral, is referred to as its polarity. Usually, a numerical score between -1 and +1 is used to represent it, with +1 denoting very positive sentiment, 0 denoting neutral sentiment, and -1 denoting very negative sentiment. Predefined lexicons (such as TextBlob and VADER) or machine learning models that link words and phrases with emotional meanings are frequently used to calculate polarity. On the one hand, using words such as excellent, happy, etc. has a positive influence on the overall score of polarity; on the other hand, words such as dreadful, sad, etc have a negative influence on the overall score of polarity. The processing of this polarity follows the tokanized document of Term Frequency (TF) of words, and it is described step-wise in Algorithm 1.

Algorithm 1: Polarity Computation from Tokenized Document

Input: Tokenized document Q

Output: Polarity score Pol (Positive, Negative, Neutral)

Step	Representation
(1) Compute importance weight (pi) for each token (qi of Q)	pi = weight(qi)

(2) Each token or sequence is input into a model	LSTM, BERT, and other transformer-based sentiment classifiers.
(3) Determine sentiment score (si) for each token using a lexicon or a model	si = sentiment(qi)
(4) Multiply each token's sentiment by its weight and sum	S = k (pi * si))
(5) Divide by the standard deviation of weights (r) to balance extremes	Pol = S/r
(6) Determine positive, negative, or neutral based on the sign of (Pol)	$\text{Polarity} = \begin{cases} \text{Positive} & Pol > 0 \\ \text{Negative} & Pol < 0 \\ \text{Neutral} & Pol = 0 \end{cases}$

- **Subjectivity:** Subjectivity is used to determine the extent of personal opinion or emotion in a text. It is a value of 0-1, with 0 being an entirely objective (neutral, fact-based, informational) statement, and 1 being an extremely subjective (biased, emotional, opinionated) statement. Unlike in the case of "The phone has an amazing display" which is a subjective statement, "The phone has a 6-inch display" is objective. These measures are explained stepwise in Algorithm 2.

Algorithm 2: Subjectivity Computation from Tokenized Document

Input: Tokenized document Q

Output: Subjectivity score Sub (Subjective or objective)

Step	Representation
1. Compute importance weight (pi) for each token (qi of Q)	pi = weight(qi)
2. Each token or sequence is input into a model	LSTM, BERT with Multi-Perspective Question Answering Subjectivity Lexicon.
3. Determine the subjectivity score (si) for each token using a lexicon or a model	si = subjectivity(qi)
4. Multiply each token's sentiment by its weight and sum	S = k (pi * si))
5. Divide by the standard deviation of weights (r) to balance extremes	Sub = S/r
6. Determine if the text is subjective or objective based on threshold T	$\text{Subjectivity} = \begin{cases} \text{Subjective} & Subj > T \\ \text{Objective} & Subj \leq T \end{cases}$

4. RESULTS AND DISCUSSION

In this paper, the experiments will start with two datasets obtained via an online Kaggle repository, and then the data preprocessing tasks will take place (as discussed in Section 3). All experiments are run on a system with a 12 th Gen Intel (R) core (TM) i7-12700H (2.30 GHz) processor, 32.0GB RAM and 64 bit windows 11 Pro operating system. Each dataset is reported separately using experiments and results to ensure that the results are well-evaluated and reliable.

4.1 DATASET-1

The experimental analysis of this dataset reveals the presence of 13 emotional labels, each with its respective sample count. The distribution, as presented in Table.4 and Fig.4, indicates a significant class imbalance. Notably, the neutral class comprises 674,538 samples, accounting for nearly 80% of the entire dataset. Such imbalance introduces challenges of data overfitting and classification bias. This is also shown in Fig.4. To mitigate these issues, we have considered all the emotions of the same count except 25% of neutral emotion samples are considered for further analysis. The Fig.5 shows some examples of (a) positive words cloud, (b) neutral words cloud, and (c) negative words cloud from Dataset-1.

Table.4. Distribution of samples of emotional classes from Dataset-1

Label	Emotion	Original count	Experimentation count
0	Anger	12,336	12,336
1	Boredom	126	126
2	Empty	5,542	5,542
3	Enthusiasm	9,304	9,304
4	Fun	10,075	10,075
5	Happiness	27,175	27,175
6	Hate	15,267	15,267
7	Love	39,553	39,553
8	Neutral	6,74,538	1,65,017
9	Relief	16,729	16,729
10	Sadness	17,481	17,481
11	Surprise	6,954	6,954
12	Worry	4,475	4,475

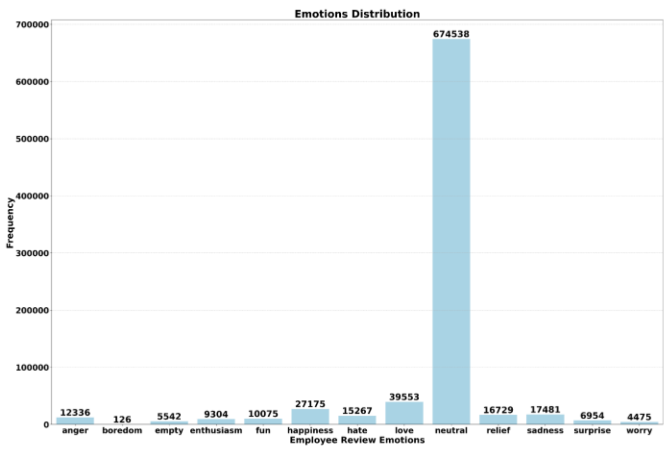


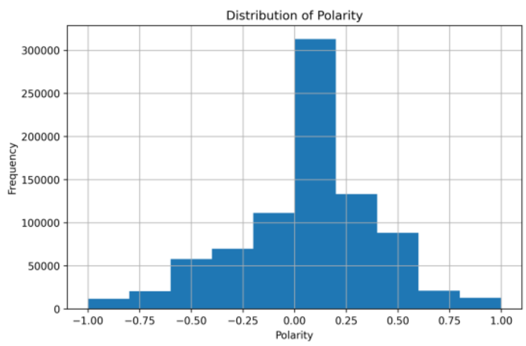
Fig.4. Distributions of 13-emotional sentiments of employee emotional comments from Dataset-1

(a)

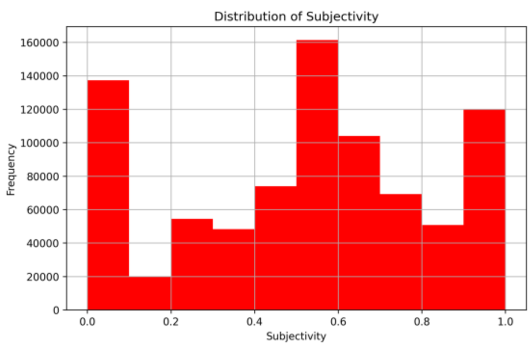
(b)

(c)

Fig.5. An example of (a) positive words cloud, (b) neutral words cloud, and (c) negative words cloud from Dataset-1



(a)



(b)

Fig.6. Distributions of polarity and subjectivity of employee emotional comments from Dataset-1

Hence, based on the study of positive, neutral and negative comments by the employees (graphs are shown in Fig.6), it has been observed that very nearer emotions to positive class are Love, Happiness, Relief. In case of negative class, the nearer comments are Hate, Sadness, and Anger. The remaining emotion classes are nearer to neutral emotions.

Table.5 demonstrates the comparative evaluation of various machine learning models based on accuracy and F1-score. Among the tested models, the Random Forest (TF-IDF) achieved the highest performance with an accuracy of 0.9856 and an F1-score of 0.9855, indicating its superior capability in handling textual data for employee review classification. The Support Vector Classifier (linear kernel, C=1.1) followed with an accuracy of 0.9451 and F1-score of 0.9385, showing strong discriminative power. The two versions of Logistic Regression (lbfgs and liblinear) gave similar performance of around 0.929 and F1-score of about 0.926, which indicates consistent and reliable performance of all solvers. Conversely, Multinomial Naïve Bayes (alpha=0.2, TF-IDF) had a relatively lower accuracy of 0.8578 and F1-score of 0.8503, which meant that it was less capable of capturing complicated patterns of textual features than the other models. In general, the model that showed the best performance in the analysis of employee reviews conducted through direct textual features is the Random Forest.

Table.5. Performance of employee review system using Dataset-1 using textual features

Model	Accuracy	F1-score
multinomialNB_alpha_0.2_tfidf	0.8578	0.8503
LogisticRegression_lbfgs_tfidf	0.9289	0.926
LogisticRegression_liblinear_tfidf	0.9292	0.9263
SVC(kernel="linear", C=1.1)	0.9451	0.9385
RandomForest_tfidf	0.9856	0.9855

Table.6. Performance of Employee sentiment System using Dataset-1 with Textual Features along with polarity and subjectivity

Model	Accuracy	F1-Score
MultinomialNB (alpha=0.2, TF-IDF)	0.867	0.86
Logistic Regression (lbfgs, TF-IDF)	0.939	0.937
Logistic Regression (liblinear, TF-IDF)	0.94	0.9375
SVC (kernel="linear", C=1.1)	0.956	0.95
Random Forest (TF-IDF)	0.997	0.9965

The Table.6 indicates the performance of Employee sentiment System on Dataset-1 with Textual Features as well as polarity and subjectivity. The results of the comparative analysis of the different machine learning models in terms of accuracy and the F1-score are presented. The Random Forest (TF-IDF) model had the best accuracy of 0.9970 and F1-score of 0.9965, which indicated that it had an exceptionally high level of performance in identifying the patterns of text and sentiment related to the review of an employee. The next best performance with a 0.9560 accuracy and an F1-score of 0.9500 came with the Support Vector Classifier (linear kernel, C=1.1) which perfectly demonstrates its strong discriminative power. The two variants of the Logistic Regression (lbfgs and liblinear) delivered similar results with an accuracy of 0.9390 and 0.9400, and F1-scores of 0.9370 and

0.9375, which shows that both optimization solvers are consistent and reliable. The Multinomial Naïve Bayes (alpha=0.2, TF-IDF) attained a comparatively lower accuracy of 0.8670 and F1-score of 0.8600, yet remained effective for basic textual classification tasks. Overall, the Random Forest model proved to be the most efficient and accurate among all models in analyzing employee reviews. These performances correspond to positive and negative classification with 13 emotions have been shown in the confusion matrix Fig.7.

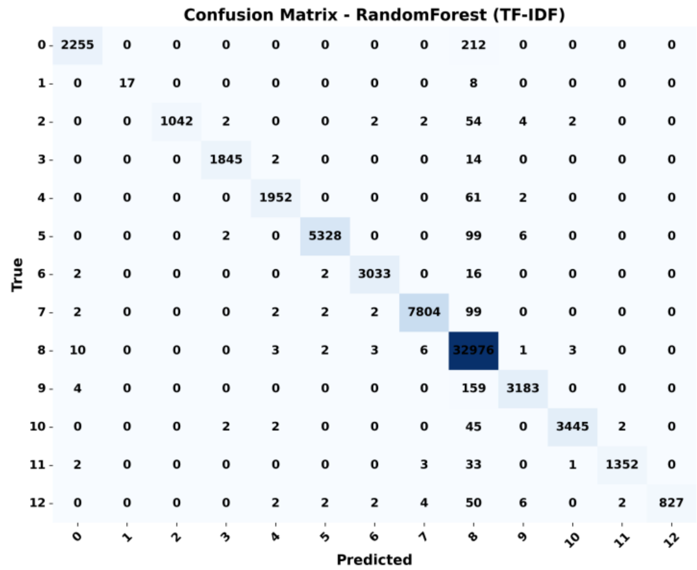


Fig.7. Confusion matrix performance of employee review system using Dataset-1

4.2 DATASET-2 EXPERIMENT

Using this dataset, the two experiments are conducted. In the first part, only 10 features ['Place', 'location', 'status', 'job_title', 'score_1', 'score_2', 'score_3', 'score_4', 'score_5', 'score_6'] (considering the data X), and ['overall'] (the label y) , have been employed to build the ESS prediction model.

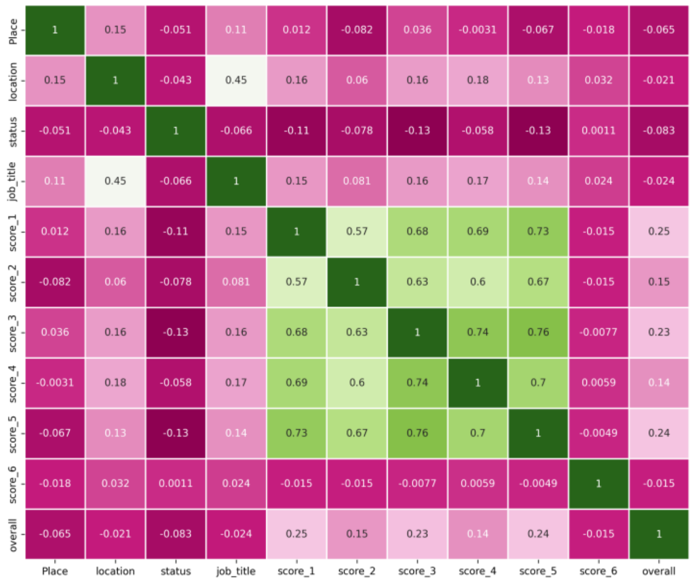


Fig.8. Correlations between features of first part experiment using Dataset-2

A correlation of these features has been shown in figure, that shows the correlation heatmap showing the linear relationships among various variables in the dataset, with correlation values ranging from –1 to +1. The green shades indicate positive correlations, pink or magenta shades indicate negative correlations, and lighter colors represent weak or no correlations. The diagonal values are all 1, as each variable is perfectly correlated with itself. Strong positive correlations are observed among the score variables (score_1 to score_5), indicating that these evaluation metrics are closely related, while the "overall" variable shows moderate correlation with these scores, suggesting it may represent an aggregate measure. In contrast, variables such as place, location, status, and job_title exhibit weak correlations with the scores, though location and job_title show a mild positive relationship. Overall, the heatmap (Fig.8) highlights that the scoring variables are internally consistent and collectively influence the overall outcome, whereas contextual factors contribute minimally.

The Table.7 presents the F1-scores of different machine learning models applied on features of Dataset 2 and configurations applied to a text classification task, highlighting how preprocessing methods and feature extraction techniques affect performance. Among all models, the Logistic Regression (liblinear) with count vectorization and cleaned text achieved the highest F1-score (0.3169), indicating it provided the best balance between precision and recall. The Multinomial Naïve Bayes (alpha = 0.2) models also performed reasonably well, especially with count vectorization and cleaned text (F1 = 0.3066), showing that simple probabilistic models can work effectively for text data. The SVC (linear and RBF) and Random Forest models yielded moderate to lower F1-scores, suggesting limited suitability for this dataset or configuration. Overall, the results show that count-based vectorization outperformed TF-IDF, and data cleaning generally improved model performance, with Logistic Regression and Multinomial Naïve Bayes emerging as the most effective classifiers for this task.

Table.7. Performance of employee review system using Dataset-2 with various classifiers

Model	Moodel_configuration	F1-score
1	multinomialMB_alpha_.2_tfidf	0.2438
2	multinomialMB_alpha_.2_tfidf	0.2438
3	multinomialMB_alpha_.2_tfidf_clean	0.2559
4	multinomialMB_alpha_.2_count	0.2967
5	multinomialMB_alpha_.2_count_clean	0.3066
6	SVC_linear_C1.1_tfidf	0.2772
7	SVC_linear_C1.1_tfidf_clean	0.2674
8	SVC_linear_C1.1_count	0.2763
9	SVC_linear_C1.1_count_clean	0.2654
10	SVC_default_rbf_C2.5_tfidf	0.2414
11	SVC_default_rbf_C2.5_tfidf_clean	0.2571
12	SVC_default_rbf_C2.5_count	0.2514
13	SVC_default_rbf_C2.5_count_clean	0.2539
14	RandomForest_tfidf	0.2049
15	RandomForest_tfidf_clean	0.1965

16	RandomForest_count	0.2291
17	RandomForest_count_clean	0.2108
18	LogisticRegression_liblinear_tfidf	0.2236
19	LogisticRegression_liblinear_tfidf_clean	0.2268
20	LogisticRegression_liblinear_count	0.3044
21	LogisticRegression_liblinear_count_clean	0.3169
22	LogisticRegression_lbfgs_tfidf	0.2412
23	LogisticRegression_lbfgs_tfidf_clean	0.2333
24	LogisticRegression_lbfgs_count	0.28
25	LogisticRegression_lbfgs_count_clean	0.2761

Influence from Table.7, in Table.8 only the numeric datatype features are considered. The prediction model has been obtained by utilizing several classifiers such as Support Vector Machine (SVM), Random Forest Classifier, AdaBoost Classifier, KNeighborsClassifier, DecisionTreeClassifier, GaussianNB, and QuadraticDiscriminantAnalysis. The performance of ERS using these classifiers has been demonstrated in Table. The Table.presents the comparative performance of different machine learning classifiers evaluated using Accuracy and F1-Score as performance metrics. Accuracy indicates the overall proportion of correctly classified instances, while the F1-Score represents the harmonic mean of precision and recall, reflecting the model's balance between correctly identifying positive cases and avoiding false positives. Among the models, the SVC, RBF kernel yielded the best results with the accuracy of 41.07% and F1-Score of 35.677 indicating that it fitted non-linear relationships in the data better than the others. RandomForestClassifier came in second with an accuracy of 38.97 and an F1-Score of 38.57 which again conforms to the results in both measures. AdaBoostClassifier and KNeighborsClassifier both had moderate performance and their accuracy was approximately 37 percent which showed that they had low predictive power. The linear SVC also had the lowest scores (36.19% accuracy and 27.58% F1-Score), indicating that such a basic linear separation could not be applied to this dataset. In general, the findings suggest that the dataset can be complicated or skewed, and non-linear classifiers such as SVC (RBF) or ensemble models are rather more efficient in revealing the underlying patterns.

Table.8. Dataset-2, which uses textual features of the employee sentiment system (ESS) performance

Classifier	Accuracy	F1-Score
SVC (RBF kernel)	41.07%	35.67%
RandomForestClassifier	38.97%	38.57%
AdaBoostClassifier	37.01%	35.14%
KNeighborsClassifier	36.87%	36.67%
SVC	36.19%	27.58%
DecisionTreeClassifier	35.74%	35.69%
GaussianNB	33.52%	29.13%
QuadraticDiscriminantAnalysis	30.45%	25.45%

In the Table.9, both numeric and textual datatype features such as ['Place', 'location', 'status', 'job_title', 'summary', 'positives', 'negatives', 'advice_to_mgmt', 'score_1', 'score_2', 'score_3', 'score_4', 'score_5', 'score_6'] transformed to ['Place', 'location', 'status', 'job_title', 'score_1', 'score_2', 'score_3', 'score_4', 'score_5', 'score_6', 'overall', 'summary_polarity', 'summary_subjectivity','positives_polarity','positives_subjectivity','negatives_polarity','negatives_subjectivity','advice_to_mgmt_polarity','advice_to_mgmt_subjectivity'] have been considered. Accordingly, the training of the classifiers uses 24268 samples composed of 15 features, and the testing of the classifiers uses 6068 samples with 15 features. The table has shown the performance of the employee review system (ERS) with actual and textual statements based features. The table provides the performance comparison of different machine learning classifiers in terms of Accuracy and F1-Score that indicate the general correctness of the classifier and the proportion between precision and recall, respectively. The best performance of all the models was recorded by the RandomForestClassifier which had an accuracy of 46.08% and an F1-score of 45.21 which means that it is a well-performing predictor with the capability to stand out due to its ability to predict the data and the complexity of the data set. The SVC (RBF kernel) also fared successfully with 45.20% accuracy and 42.46% F1-score indicating that the non-linear learning via kernel methodology was able to locate valuable data patterns. AdaBoostClassifier and KNeighborsClassifier performed fairly well with moderate results (approximately 41 percent) which is good but has weak generalization. The scores of the linear SVC, DecisionTreeClassifier, GaussianNB and QuadraticDiscriminantAnalysis were lower, and therefore, it could be inferred that the simple or linear models could not capture the underlying data relationships. Overall, the results indicate that ensemble-based and non-linear classifiers, particularly Random Forest and SVC (RBF), demonstrate superior performance for this task, likely due to their ability to handle non-linear dependencies and feature interactions more effectively.

Table.9. Performance of Employee Review System (ERS) with textual Features along with Polarity & Subjectivity

Classifier	Accuracy	F1-Score
SVC (RBF kernel)	45.20%	42.46%
RandomForestClassifier	46.08%	45.21%
AdaBoostClassifier	41.18%	40.79%
KNeighborsClassifier	40.98%	35.80%
SVC	40.01%	32.61%
DecisionTreeClassifier	38.56%	37.53%
GaussianNB	37.42%	35.33%
QuadraticDiscriminantAnalysis	38.02%	33.71%

Since the test samples lack actual class labels or ground truth, the predicted labels generated before applying polarity and those obtained after applying polarity are compared using Normalized Mutual Information (NMI). The NMI score provides a measure of similarity between the two sets of label assignments. My NMI score of 0.2015 between Model A and Model B predictions indicates very low similarity in their label assignments. In other words, the two models are grouping/classifying the test samples in quite different ways. However, this value does not reveal which model performs better, since ground-truth labels are not available, it only shows that the models strongly disagree with each other.

Both models exhibit low agreement with the natural clustering structure of the test data, as reflected in their small NMI and ARI values, which indicates that the predicted labels from neither model perfectly capture the inherent grouping patterns within the feature space. However, Model A (prediction model after polarity) achieves consistently higher scores than Model B (prediction model before polarity) (NMI: 0.168 vs. 0.137, ARI: 0.090 vs. 0.062). Since NMI and ARI measure the similarity between predicted labels and data-driven clusters, higher values imply that the predictions are better aligned with the underlying structure of the data. Therefore, even though the absolute values are low, the relative difference justifies considering Model A as more reliable than Model B in this context, especially given the lack of ground truth labels for direct accuracy evaluation.

Hence from the above experiment, it has been observed that the employee sentiment analysis system works well on textual data along with textual polarity and subjectivity. Even if there is possibility of extracting polarity and subjectivity addition with numeric features in absence of textual features.

5. CONCLUSIONS

This work represents a method of employee sentiment analysis and their attitudes towards an organization by uncovering the employee sustainability in the organization. This work establishes a framework of electronic human resource management activities with sentiment activity, and for this the work proposes an employee sentiment analysis system where both textual and non-textual content of the employee have been considered. Then data driven to data analysis tasks have been performed to build the prediction model. Finally, the prediction model has been validated with two standard datasets that finds the cause effect is the usability of textual and non-textual information of employees for uncovering their sentiments about the organizations.

REFERENCES

[1] S. Bird, E. Klein and E. Loper, "*Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*", O'Reilly Media, 2009.

[2] T. Bondarouk and C. Brewster, "Conceptualising the Future of HRM and Technology Research", *The International Journal of Human Resource Management*, Vol. 27, No. 21, pp. 2652-2671, 2026.

[3] A. Geron, "*Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*", O'Reilly Media, 2022.

[4] D. Hirlea and M. Rei, "Contextual Sentence Classification: Detecting Sustainability Initiatives in Company Reports", *Proceedings of International Conference on Machine Learning and Deep Learning*, pp. 1-8, 2021.

[5] P. Jain and M. Singh, "Human Resource Management Practices for Sustainable Organizational Development", *International Journal of Organizational Analysis*, Vol. 30, No. 6, pp. 1411-1426, 2022.

[6] D.T. Kooij, M. Van Woerkom and J.J. Denissen, "Job Crafting Towards Strengths and Interests: The effects of a Job Crafting Intervention on Person Job Fit and the Role of Age", *Journal of Applied Psychology*, Vol. 105, No. 4, pp. 338-352, 2020.

[7] F.S.U. Putra and I. Puspitasari, "Implementation of Sentiment Analysis in HR Management and Development Planning using Topic Modelling on Employee Review", *Kontigensi: Jurnal Ilmiah Manajemen*, Vol. 11, No. 2, pp. 409-417, 2023.

[8] L. Rink, J. Meijdam and D. Graus, "Aspect-Based Sentiment Analysis for Open-Ended HR Survey Responses", *Proceedings of International Conference on Management Studies*, pp. 1-8, 2024.

[9] K. Stachova, Z. Stacho and F. Sekan, "The Impact of E-HRM Tools on Employee Engagement", *Administrative Sciences*, Vol. 14, No. 11, pp. 303-315, 2024.

[10] S. Strohmeier, "Digital Human Resource Management: A Conceptual Clarification", *German Journal of Human Resource Management*, Vol. 34, No. 3, pp. 345-365, 2020.

[11] S. Strohmeier and F. Piazza, "Artificial Intelligence Techniques in Human Resource Management-A Conceptual Exploration", *Intelligent Systems in Accounting, Finance and Management*, Vol. 22, No. 4, pp. 273-286, 2015.