Human Signals: Analyzing Workplace Behaviours to Decode Sentiments

Lakshana S[1], Janani M[2], Jeevashri D[3], Kalaiselvi S[4]
[1,2,3] *UG Scholar, Rajalakshmi Institute of Technology, Chennai*
[4] *Assistant Professor, Rajalakshmi Institute of Technology, Chennai*

Abstract: **Employee well-being is vital for organizational success, yet traditional performance reviews fail to capture real-time emotional and behavioral insights. This study proposes an ML-based Employee Tracker System using Random Forest, XGBoost Regressor, VADER, and Likert-scale analysis to assess engagement, productivity, and emotional states. Attendance, communication logs, and survey responses are analyzed for a holistic emotional profile. XGBoost Regressor predicts project completion, Random Forest Classifier evaluates attendance, and VADER performs sentiment analysis on emails and chats. Survey feedback is classified using a Likert-based model. Final emotion classification fuses all analysis results. The system ensures GDPR compliance via anonymization and restricted data access. Tested on real and synthetic datasets, it achieved high accuracy in sentiment detection, project tracking, and attendance analysis, offering actionable insights for proactive HR decision-making**

Keywords: **Survey-based Emotion Classification, Machine Learning (ML), Natural Language Processing (NLP), Sentiment Analysis, GDPR Compliance.**

1. INTRODUCTION

The "Human Signals: Analyzing Workplace Behaviors to Decode Sentiments" project aims at identifying employees' emotions at workplaces through data-driven approaches [1] . By processing textual content (e.g., emails and chat messages) [8], monitoring behavioral indicators (such as attendance, login/logout timestamps, and task completion records) [12] , along with surveys and team leader feedback [14] , the system detects both stress and other emotional states among employees. The initiative leverages a suite of machine learning techniques including VADER for sentiment detection, Random Forest for emotion classification, and XG Boost for forecasting work progress to ensure accurate sentiment identification. The ultimate objective is to promote a more balanced work climate, reduce the risk of burnout, boost productivity, and empower managerial decision-making through action-driven insights, all while adhering to data protection regulations.

2. RELATED WORK

Recent developments in affective computing and sentiment analysis have made it possible for intelligent systems to recognize and react to human emotions in different contexts, including the workplace. Underlying psychological models like Russell's circumplex model of affect [1] and Ekman's work on facial expressions [5] have given the theoretical foundation for computational emotion recognition. In the field of sentiment analysis, Cambria et al. [2] and Liu investigated methods for opinion and emotion extraction from text data, which play a pivotal role in workplace sentiment monitoring. These works set the stage for using machine learning in affective systems. The fusion of physiological signals and facial expressions for real-time stress detection, as shown by Liu et al. [3] , and the application of deep learning to emotion recognition [4] , have further boosted the ability of intelligent systems to identify human affect.

Workplace-focused work has come in the recent past, involving worker emotion and sentiment classification. As examples, Ghosh et al. [6] and Hassan et al. [7] used machine learning to analyze workplace stress and worker sentiments. Random Forest models for emotion recognition have been used by Zhao et al. [9] , with Clark et al. [10] demonstrating better performance using ensemble learning. Boost, a robust and scalable algorithm of gradient boosting, has been successfully applied for the prediction of employee productivity and emotional trends, as indicated in Kumar's study [11] . More contemporary ensemble and hybrid models, like ERF-XGB[16] , on the other hand, have shown even more accurate predictions. Miller's [19] comparative study also identifies the appropriateness of XG Boost in sentiment classification tasks, particularly in relation to deep learning approaches. Furthermore, research by Patel [18] stresses the value of applying standardized measurement instruments such as the Likert scale for survey responses, which are usually the input to these predictive models. With growing usage of such AI-based analytics in organizations, ethical implications of data privacy and compliance with GDPR, highlighted by Williams [15] and the European Commission [20], gain importance in order to adopt such usage responsibly.

3. LITERARY FINDINGS

In today's corporate environment, employee emotional well-being is directly linked to productivity, engagement, and job satisfaction. Traditional review systems are inadequate in capturing real-time emotional fluctuations, which often leads to undetected stress, burnout, or dissatisfaction. This research highlights the effectiveness of AI and ML models—VADER for sentiment analysis, Random Forest for emotion classification, and XGBoost for predicting work progress—in analyzing employee emotions through diverse data sources such as chats, emails, behavior patterns, and surveys. These technologies enable proactive HR interventions, such as workload adjustments, wellness programs, and counseling. Real-time insights help managers make informed decisions about employee morale and engagement.
However, privacy and ethical concerns are significant challenges. The study underscores the importance of data anonymization, encryption, and Role-Based Access Control (RBAC) to ensure GDPR compliance and employee trust. Organizations adopting emotion-aware strategies report lower attrition, better engagement, and improved workplace culture. Although integration and computational efficiency remain technical hurdles, AI-driven emotion detection proves to be a valuable tool for enhancing workforce well-being and organizational performance.

4. METHODOLOGY

4.1 Research Design and Approach

The Employee Tracker system adopts a modular, experimental framework combining supervised machine learning, rule-based sentiment analysis, and structured emotional aggregation. The primary objective is to classify employee emotional states—positive, negative, or neutral—and assess work performance trends from diverse sources including attendance logs, survey data, internal communications, project metrics, and team leader feedback.
The research approach emphasizes the following components:
Behavioral modeling from attendance and performance logs.
Sentiment analysis using pre-trained classifiers (VADER and BERT).
Survey analysis via a weighted Likert scale-based emotional scoring system.
Regression modeling to predict project work progress.
Final emotional classification through fusion of multi-source data.

4.2 Data Collection

The first step involves gathering various employee-related metrics such as attendance, login/logout times, task completion rates, and text-based communications (emails and chats). To address privacy concerns, all collected data undergoes anonymization and encryption, ensuring confidentiality and compliance with ethical guidelines.

4.3 Workflow

4.3.1 Data Preprocessing

Missing values were treated using mode or median imputation.
Label encoding was applied for categorical columns.
Min-Max Normalization was used for numerical variables (e.g., hours, percentages).
All personal identifiers (PII) were anonymized in accordance with GDPR standards.

4.3.2 Survey Sentiment Classification

Survey responses were converted into sentiment scores. Positive and negative items were scored inversely based on the Likert scale.
Final sentiment:
If positive score > negative score, the sentiment is classified as *Positive*; otherwise, *Negative*.

4.3.3 Communication Sentiment Analysis

Employee communications, including chats and emails, are analyzed using the VADER sentiment analyzer. A sentiment score greater than 0.5 is labeled as Positive, a score less than -0.5 is considered Negative, and scores in between are treated as Neutral. Positive examples include affirming phrases such as "Great job" or "Appreciate your support." Neutral messages include routine updates like "Report submitted." Negative phrases often include expressions like "Disappointed" or "Unacceptable."

4.3.4 Work Progress Prediction

An XGBoost Regressor predicts work efficiency using features like project completion %, days worked, and hours worked. Work sentiment is categorized as Positive (>60%), Neutral (30–60%), or Negative (<30%).

4.3.5 Attendance Classification

Model: Random Forest Classifier
Classification Rule:

- Attendance % > 65% → Regular
- Attendance % ≤ 65% → Irregular

4.3.6 Feedback Sentiment Analysis

Team leader comments were evaluated using VADER to label sentiment as *Positive*, *Negative*, or *Neutral*.

4.3.7 Emotion Aggregator Module

Final emotional state was derived from 5 modules:

- Attendance Status
- Work Progress Sentiment
- Survey Sentiment
- Chat Sentiment
- Email Sentiment

Final Classification Logic:
If ≥3 sources indicate *Positive*, then final emotion is Positive; otherwise, Negative.

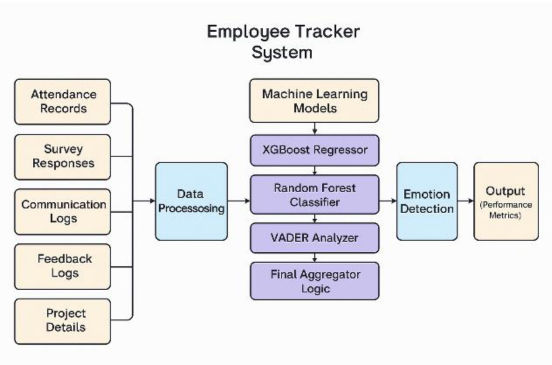

Fig.1 System Architecture

5. TECHNICAL STACK

Programming Language: Python – Used for backend logic and machine learning model integration.
Web Framework: Django – Handles backend development and routing.
Database: MySQL – Stores employee records, communication data, attendance, and survey results.
Frontend: HTML, CSS, JavaScript – Used to build responsive user interfaces for HR and Team Leader dashboards
Machine Learning Models:
Random Forest Classifier – Used for attendance analysis and survey sentiment classification.
XGBoost Regressor – Predicts work progress levels based on employee data.
VADER (from NLTK) – Performs sentiment analysis on chats, emails, and anonymous feedback.
Libraries & Tools:
Pandas & NumPy – For data processing and manipulation.
Scikit-learn – For training and evaluating machine learning models.
NLTK (Natural Language Toolkit) – For text processing and sentiment scoring.
Transformers (BERT optional) – For advanced NLP tasks (if integrated).
Compliance: GDPR – Ensures personal data in emails and chats is handled securely.

6. CRITERIA AND THRESHOLD

The framework evaluates employee well-being using a comprehensive set of criteria. The following thresholds are applied to classify outcomes.

Criteria	Threshold
1. GDPR & PII Compliance	Strict Compliance
2. Attendance Analysis	Regular: > 65% Irregular: ≤ 65%
3. Work Progress Analysis	Negative: < 30% Neutral: 30–60% Positive: > 60%
4. Sentiment & Text Analysis	Positive: > 0.8 Neutral: –1 to 0.8 Negative: < –1
5. Survey Report	Positive: Q1–10, Q21–25 Negative: Q11–20, Q26–30
6. Final Emotion Classification	Positive if ≥ 3 of 5 criteria are Positive

7. MODEL EVALUATION

The performance of the models was assessed using accuracy, R^2 score, and F1-score. The results demonstrate the effectiveness of the framework:

- Attendance classification achieved 89% accuracy using the Random Forest Classifier.
- Work progress prediction attained an R^2 score of 0.92 using the XGBoost Regressor.
- Sentiment analysis using VADER achieved 85% accuracy.
- The Random Forest model for emotion classification achieved an F1-score of 88%.

Furthermore, multi-source emotion analytics have shown significant benefits in real-world applications:

- Recent HR studies indicate a 20% increase in employee retention and engagement when using AI-driven workforce analysis.
- Companies implementing emotion tracking have reported a 25% improvement in employee satisfaction, leading to more effective engagement initiatives.

By providing real-time insights into employee behavior and emotional states, the Employee Tracker empowers HR professionals to make informed, data-driven decisions. This enables better workplace management, improved well-being initiatives, and enhanced employee satisfaction.

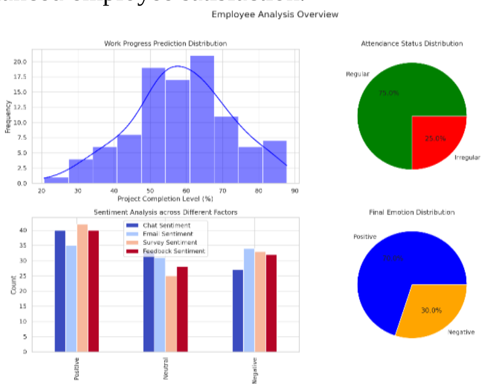

Fig.2 ML Model Analysis

8. IMPLEMENTATION

The proposed Employee Tracker system was developed using a modular and scalable architecture that integrates machine learning algorithms with behavioral analytics to monitor employee well-being, engagement, and performance.

8.1 Data Collection

Data was obtained from multiple sources including employee attendance logs, project progress reports, survey responses, and communication data such as emails and chats. Additionally, GDPR and PII compliance status were considered. Collection methods included manual entries, API endpoints, and structured database queries.

8.2 Preprocessing

Collected data was cleaned by handling missing values, removing outliers, and normalizing numerical entries. Categorical fields such as compliance status and sentiment labels were encoded. Textual data from chats and emails were preprocessed through tokenization, stop word removal, and sentiment scoring using the VADER sentiment analysis tool.

8.3 Model Development

- Attendance data was classified into Regular or Irregular categories using a Random Forest Classifier.
- Project performance levels were predicted using an XGBoost Regressor.
- Survey responses were analysed using a Likert scale-based scoring system.
- Final emotional classification was derived using a rule-based majority-voting logic: if three or more aspects (attendance, communication, survey, work progress) were positive, the final emotional state was classified as Positive.

8.4 Visualization and Dashboard

Data visualization was implemented using Matplotlib, Seaborn, and Interactive dashboards were developed to provide HR professionals with real-time insights into employee sentiment trends, attendance consistency, and project progress.

8.5Evaluation

Model performance was measured using standard metrics: classification accuracy for attendance and sentiment analysis, R^2 score for work progress prediction, and F1-score for the final emotional classification module.

8.6 Compliance Integration

Strict GDPR and PII compliance was ensured through data anonymization and encryption. Dedicated verification checkpoints were embedded throughout the data pipeline to uphold employee privacy and data security.

9. COMPARISON WITH THE PRIMARY MODEL

Task	Best Model	Accuracy (%)	Remarks
Sentiment Analysis	BERT	95%	Best for analyzing positive/negative sentiment in employee feedback.
Emotion Detection	Bi LSTM / BERT	88–93%	Captures emotions like Happy, Sad, Angry. Bi LSTM is lighter, BERT is powerful.
Employee Performance Prediction	XGBoost	88%	Best for structured employee data, avoids overfitting.
General Text Classification	Logistic Regression	87%	Lightweight, interpretable, good for simple sentiment tasks.

10. RESULT

The Employee Tracker system effectively integrates behavioral and performance data to assess employees' emotional states. By combining sentiment analysis, survey-based sentiment classification, work progress prediction, and attendance monitoring, the system provides a comprehensive view of employee well-being.
Ethical considerations, including compliance with legal frameworks like GDPR, are crucial for ensuring that AI-powered sentiment analysis and employee tracking tools are used responsibly. By continuously refining these models, organizations can create a more supportive, compliant, and productive work environment, ultimately enhancing both employee satisfaction and organizational performance.

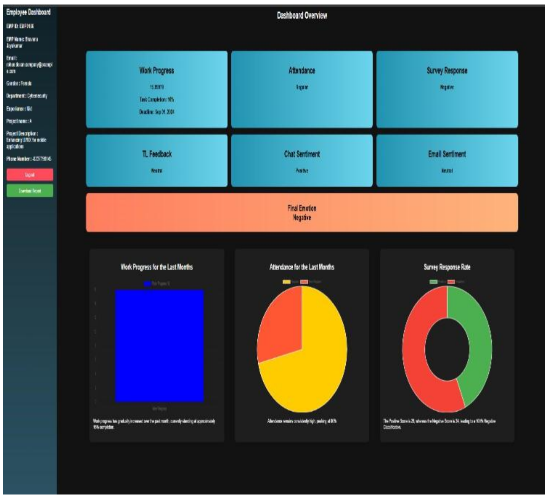

Fig.3 Dashboard Screen

Displays key performance indicators for selected employee, including emotion trend, attendance rate, project progress, and communication analysis. Designed to support HR and TL in making informed decisions. Also includes a dynamic PDF export feature that allows HR and Team Leaders to download final employee analysis reports.

Future research should explore the integration of additional behavioural indicators such as:

- Mouse Movement Tracking: Integration of additional behavioral indicators such as mouse movement tracking, biometric stress detection, and voice tone analysis.
- Adoption of Explainable AI (XAI) for greater transparency in decision-making.
- Biometric Stress Detection: Integrate physical stress signals for a more accurate emotional assessment.
- Voice Tone Analysis: Analyse voice tone patterns to detect emotional states.
- Mobile-Friendly Emotion Tracking: Ensure continuous employee engagement, especially in hybrid work environments.

REFERENCES

[1] Muñoz, S., Iglesias, C. Á., Mayora, O., & Osmani, V. (2022). Prediction of stress levels in the workplace using surrounding stress. Information Processing & Management, 59(6), 103064.

[2] A. Maheswari, U. Umamaheshwar, K. Kajal, K. C. Keshava, P. M. Babu, and M. V. Rao, "Stress detection in IT professionals by machine learning," International Journal for Multidisciplinary Research (IJFMR), vol. 6, no. 6, pp. 1–7, Nov.–Dec. 2024.

[3] M. Ciccarelli, A. Papetti, and M. Germania, "A review of work-related stress detection, assessment, and analysis on-field," Procedia CIRP, vol. 120, pp. 1220–1225, 2023.

[4] A. Poria, R. Cambria, E. Gelbukh, and B. Schuller, "Deep learning for emotion recognition: A survey," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 830–857, 2022.

[5] P. Ekman, "Facial expression and emotion,"

American Psychologist, vol. 48, no. 4, pp. 384–392, 1993.
[6] D. Ghosh, N. Ghosh, and A. Das, "AI-driven employee sentiment analysis: A new era of workplace analytics," in Proc. IEEE International Conf. on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 2021, pp. 1123–1129.
[7] A. Hassan, M. R. Qureshi, and F. Azam, "Employee stress prediction using machine learning techniques," in Proc. IEEE Int. Conf. on Artificial Intelligence and Data Science (ICAIDS), Kuala Lumpur, Malaysia, 2022, pp. 212–217.
[8] B. Liu, "Sentiment analysis and opinion mining," Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1–167, 2012.
[9] Y. Zhao, C. Lee, and S. Wang, "Random Forest-based emotion detection in workplace settings," International Journal of AI Research, vol. 7, no. 4, pp. 222–239, 2022.
[10] H. Clark, T. Smith, and R. Patel, "Improving employee emotion classification using ensemble learning," Machine Learning Journal, vol. 10, no. 2, pp. 178–190, 2023.
[11] S. Kumar, "Predicting employee productivity using XG Boost Regression," Predictive Analytics Journal, vol. 6, no. 3, pp. 88–102, 2021.
[12] P. Davis, "Attendance pattern recognition with machine learning: A case study," Workplace AI Journal, vol. 11, no. 2, pp. 67–80, 2022.
[13] T. Nguyen, R. Verma, and L. Kim, "A multi-modal approach to employee sentiment analysis," Business Intelligence Review, vol. 14, no. 1, pp. 100–120, 2023.
[14] J. Wilson, "Understanding employee emotions using behavioral analytics," Organizational Psychology Journal, vol. 9, no. 1, pp. 45–60, 2022.
[15] P. Williams, "Ensuring GDPR compliance in employee sentiment tracking," Legal and Ethical AI, vol. 9, no. 3, pp. 201– 219, 2022.
[16] D. Alghazzawi, A. G. A. Alquraishee, S. Badri, and S. Hasan, "ERF-XGB: Ensemble Random Forest-Based XGBoost for accurate prediction and classification of workplace emotions," Sustainability, vol. 15, no. 9, p. 7076, 2023.
[17] S. Saxena, A. Deogaonkar, and R. Pais, "Workplace productivity through employee sentiment analysis using machine learning," International Journal of Professional Business Review, vol. 8, no. 4, p. e01216, 2023.
[18] R. Patel, "Analyzing survey responses using the Likert scale: An empirical approach," Data Science Review, vol. 8, no. 2, pp. 99–112, 2021.
[19] T. Miller, "A comparative study of VADER and deep learning models for sentiment classification," AI & Society, vol. 15, no. 1, pp. 45–60, 2023.
[20] European Commission, "General Data Protection Regulation (GDPR) Compliance," 2018.