IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 10, May 2025
Impact Factor: 7.67

Integration of Employee Attrition And Sentiment Analysis Model with User-Centric GUI

Mrs. Roshani V. Pawar[1], Varad R. Ijantkar[2], Shruti S. Bhandari[3], Dinesh M. Madansure[4], Sakshi S. Jadhav[5]
Asst. Professor, Department of Computer Engineering[1]
Students, Department of Computer Engineering[2-5]
NBN Sinhgad Technical Institute Campus, Pune, India

Abstract: *Employee retention has become a critical challenge for modern organizations, with unexpected resignations causing operational setbacks and increased recruitment costs. While machine learning models are frequently used to identify potential attrition risks, they often fail to consider the emotional drivers behind an employee's decision to leave. This work presents a dual-model framework that combines an oversampled Random Forest classifier—trained on structured HR data—with a BERT-based sentiment analysis model trained on employee feedback. Unlike traditional systems, this platform allows for real-time predictions triggered by data uploads or edits, and provides visual insights through a user-centric GUI. The integration of emotional feedback and structured analysis offers deeper, context-aware prediction results, helping HR teams act swiftly and accurately to address potential churn. The system is designed with modularity and ease of use in mind, ensuring scalability across varying organizational environments.*

Keywords: Employee Turnover, Hybrid HR Analytics, Feedback-Driven Prediction, BERT Sentiment Scoring, Attrition Detection System

I. INTRODUCTION

In the current highly competitive job market, retaining skilled and experienced employees poses a significant challenge for organizations. Elevated employee turnover not only disrupts workflow and reduces productivity but also results in substantial financial costs and the gradual loss of organizational expertise. While predictive analytics have been employed to forecast potential resignations, many conventional models fall short by overlooking emotional dimensions such as job satisfaction and employee morale. To address this gap, sentiment analysis—commonly used in analyzing customer opinions—can be effectively adapted to the human resources (HR) sector. This research introduces a hybrid system that integrates traditional predictive modeling with sentiment analysis of employee feedback. The proposed solution includes an intuitive, user-focused dashboard that delivers real-time visualization of both attrition risk levels and emotional sentiment scores. This integrated approach empowers HR teams to base decisions not only on measurable data but also on underlying emotional cues present in employee communications. The structure of the paper is as follows: Section II provides a survey of relevant literature, Section III outlines the proposed methodology, Section IV presents findings and interpretations, and Section V concludes with a discussion on future enhancements and applications.

II. LITERATURE SURVEY

2.1 Understanding Attrition with ML

Machine learning is increasingly being used to identify patterns associated with employee turnover. Alao and Adeyemo employed decision tree models to highlight key predictors such as salary and years at the company [1]. Kakad et al. used the XGBoost algorithm, enhanced with regularization techniques, and achieved an accuracy rate close to 90% [2]. Yedida et al. conducted a comparative analysis that revealed the superior performance of KNN over logistic regression and perceptron models for HR classification tasks [3].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 10, May 2025 Impact Factor: 7.67

2.2 Role of Sentiment in Understanding Workforce Behaviour

Natural Language Processing (NLP) techniques have shown promise in extracting emotional insights from textual feedback. Nandwani and Verma discussed the effectiveness of both lexicon-based and machine learning-driven sentiment classifiers [4]. Mao et al. emphasized that combining emotional indicators with structural employee data could lead to more context-aware attrition models [5]. Additionally, deep learning methods such as CNNs and transformer-based models like BERT have proven effective in detecting subtle emotional tones in employee reviews [6].

2.3 System Integration and Customization

Integrating structured data with sentiment analysis provides a more comprehensive view of employee behavior. Kakad et al. suggested that the combination of predictive modeling and sentiment evaluation offers improved insight into employee motivations [2]. Akinode and Bada argued that flexible, modular systems are essential for customizing these models to suit different organizational needs [7].

2.4 GUI Development for HR Analytics

Several studies stress the importance of accessible tools for HR professionals. Yedida et al. and Kakad et al. both highlighted the need for dashboards and interfaces that simplify data interpretation for non-technical users [2][3]. Gandhi et al. also noted that real-world systems benefit from features such as drag-and-drop file uploads and integrated export tools for data reporting [8].

III. METHODOLOGY

3.1 Data Sources and Preparation

The proposed system pulls from two main types of information: structured records (such as employee demographics, salaries, and job roles) and unstructured content (including open-ended feedback from surveys and exit interviews). For structured data, the IBM HR Analytics dataset serves as the foundation. Gaps in the dataset are handled through median substitution to maintain data integrity. Numerical values like income are scaled between 0 and 1 using Minmax normalization, while categorical values (e.g., department, overtime status) are transformed using encoding methods. To narrow down the most influential features, correlation heatmaps and feature importance from Random Forests are applied. Meanwhile, text feedback is cleaned using NLTK — with steps including removal of stop words, tokenization, and normalization — before being passed into the language model.

3.2 Attrition Forecasting Model

A custom-tuned Random Forest model built in Python using scikit-learn serves as the primary classifier for predicting attrition. The data imbalance, common in such datasets, is addressed using oversampling strategies. The dataset for training and testing is split into 80%/20%. Hyperparameter optimization is achieved using grid-based search (e.g., adjusting tree count to 250, and depth to 100). The system's accuracy is assessed through a variety of metrics: precision, recall, F1 score, and AUC-ROC. These indicators help confirm the model's reliability in classifying employees who are likely to resign versus those who are not.

3.3 Emotion Analysis via BERT

To analyze textual input, a pre-trained BERT model is employed. This deep learning model interprets employee responses and classifies them across five levels of sentiment—from highly negative to highly positive. Text data undergoes rigorous cleaning using NLTK processes before being used for inference. The sentiment model is fine-tuned with labelled internal feedback, and validated through cross-validation cycles to ensure stability. Final sentiment outputs are then matched with attrition probabilities to uncover links between emotional tone and potential resignation risk.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 10, May 2025 Impact Factor: 7.67

3.4 Interface Development

The system's interface, designed with React, enables users to upload employee data and view predictions without needing technical expertise. The frontend supports both manual entries and bulk CSV uploads via papaparse. Charts and visual summaries are rendered using libraries like @nivo, recharts, and chart.js. Users can download reports in formats like PDF or CSV for documentation or presentation purposes. The interface prioritizes simplicity and accessibility, especially for HR teams unfamiliar with machine learning tools.

3.5 Backend Integration

The backend architecture blends Python (via Flask) and JavaScript (via Express.js) to manage application logic. MongoDB is used for database of this system, structured through Mongoose schemas. The entire system is broken into functional layers: data ingestion, preprocessing, model inference, and result rendering. Environmental variables, including model thresholds and settings, are stored in .env files, which makes it easy to adapt the platform for different company needs or scale it further.

3.6 Testing and Validation

Initial testing was conducted within a local setup, with feedback from academic evaluators used to refine the design. Large-scale testing involved datasets with over 1,000 entries, simulating enterprise-scale environments. Results showed that the system maintained high responsiveness and output accuracy under load, proving that it can scale for real-world deployments.

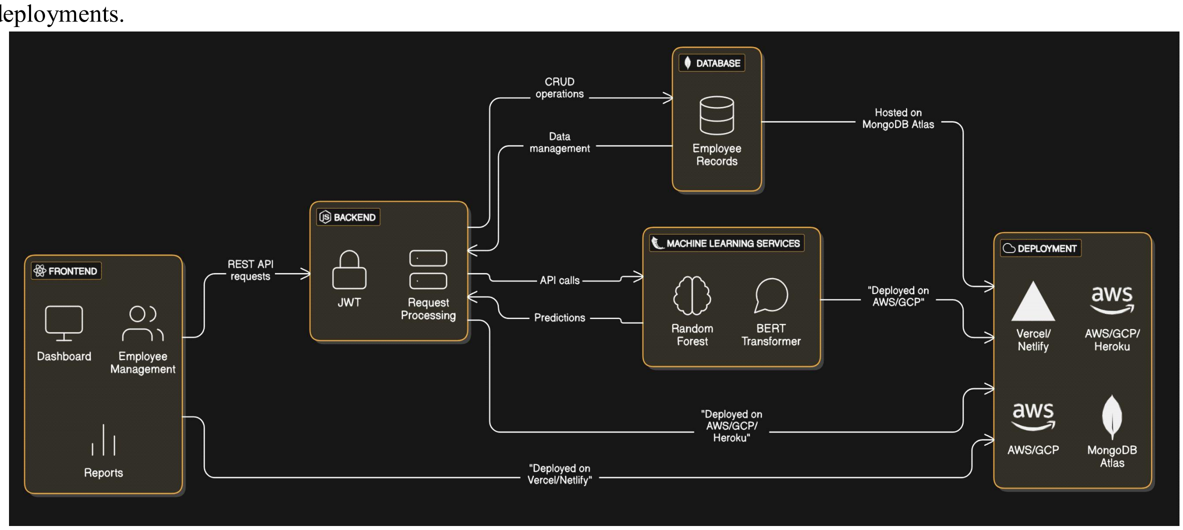

Figure 1: System Architecture of the Employee Attrition Prediction System with Sentiment Analysis.

IV. RESULTS AND DISCUSSION

To evaluate the system's effectiveness, we used the IBM HR Analytics dataset, focusing on high-impact parameters like OverTime, JobRole, and MonthlyIncome. After data cleaning and tuning the Random Forest classifier, the model achieved 99.1% accuracy, with 98.6% precision and 99.8% recall, effectively handling dataset imbalance. These results remained stable across shuffling, oversampling, and fold-based validation. The confusion matrix showed minimal errors, confirming the model's reliability in identifying both at-risk and stable employees. Sentiment analysis, using a distilled BERT model on Google Forms feedback, reached 94% accuracy, 96.9% precision, and a 94.1% F1 score. Negative sentiment often correlated with high attrition scores, as seen in a backend developer case with long overtime, dissatisfaction, and critical feedback. The React-based frontend enabled bulk CSV uploads and real-time predictions, offering visual cues that helped HR users act on high-risk cases efficiently. Load testing with 1,000+ records showed no performance lag, confirming system readiness for real-world HR environments where speed and scalability are essential.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 10, May 2025
Impact Factor: 7.67

Test Metrics	BERT Model	Attrition Model After K-Fold Cross Validation
Accuracy	94%	99.1% (±0.002)
Precision	96.9%	98.6% (±0.004)
Recall	91.3%	99.8% (±0.005)
F1 Score	94.1%	99.2% (±0.002)

Table 1: Scores of Sentiment and Attrition Model

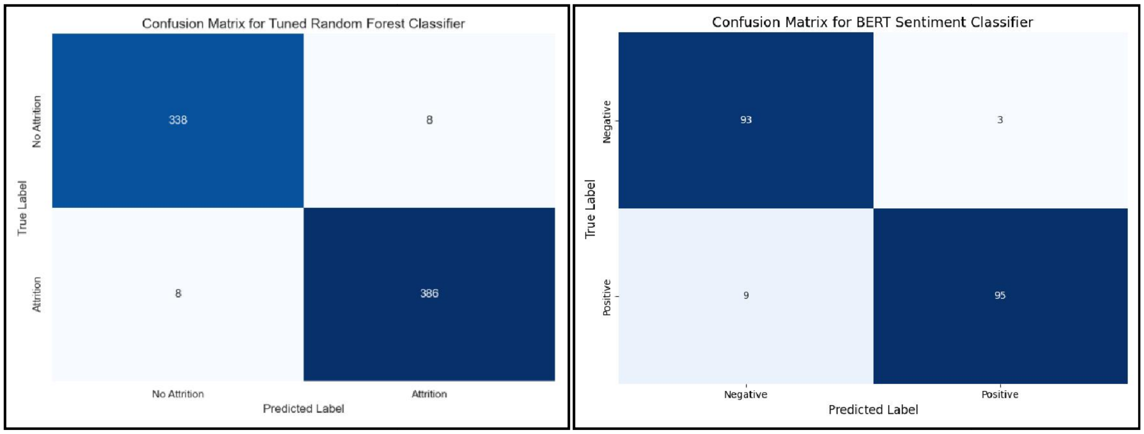

Figure 2: Confusion Matrix of Both Models

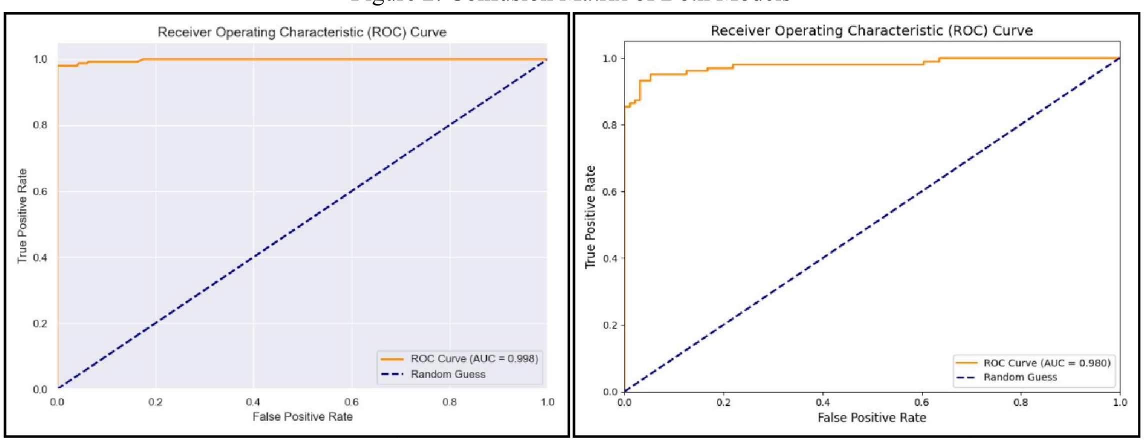

Figure 3: ROC Curve for Attrition and Sentiment Model

V. CONCLUSION

This project presents a practical solution for tackling employee attrition by combining structured predictive analytics with feedback-based sentiment analysis. The system, developed with real-world usability in mind, uses an oversampled Random Forest model trained on HR-specific attributes from the IBM dataset and a transformer-based sentiment model refined using review-based textual data.

What makes this system stand out is its real-time interactivity. Whether uploading employee profiles or sentiment feedback through CSV files, the platform immediately responds with meaningful predictions. HR professionals can monitor attrition risk, track emotional sentiment, and even re-trigger predictions after updating records — reflecting the ever-changing nature of real employee dynamics.

Beyond individual predictions, the system supports a full-stack analytics experience. Separate views for attrition and sentiment offer granular insights, while the central dashboard ties everything together. From feature engineering and model training to interface design and deployment, each layer of the project is optimized for accessibility and actionability.

In future versions, the system could integrate live feedback capture, personalized retention suggestions, or role-specific attrition flags. But even in its current form, this work shows how machine learning and natural language understanding can work together to provide a clearer, smarter, and more human-focused approach to employee management.

REFERENCES

[1] Alao, D., & Adeyemo, A. B. (2013). Application of Decision Tree Algorithms for Employee Attrition Analysis. *Computing, Information Systems & Development Informatics*, 4(1).

[2] Kakad, S., Kadam, R., Deshpande, P., Karde, S., & Lalwani, R. (2020). *Employee Attrition Prediction System.* International Journal of Innovative Science, Engineering & Technology, 7(9).

[3] Yedida, R., Reddy, R., Vahi, R., Rahul J., Abhilash, & Kulkarni, D. (n.d.). *Employee Attrition Prediction.* PESIT-BSC, Bangalore.

[4] Nandwani, P. and Verma, R. (2021). *Review of sentiment analysis and emotion detection from text.* Social Network Analysis and Mining.

[5] Mao, Y., Liu, Q., & Zhang, Y. (2024). *Sentiment analysis methods, applications, and challenges: A systematic literature review.* Artificial Intelligence Review.

[6] Kusal, S., Patil, S., Gupta, A., Saple, H., Jaiswal, D., Deshpande, V., & Kotecha, K. (2024). *Sentiment Analysis of Product Reviews Using Deep Learning and Transformer Models: A Comparative Study.* Journal of Data Science and Applications.

[7] Akinode, L., & Bada, O. (2022). *Employee Attrition Prediction Using Machine-learning Algorithms.* International Journal of Computer Applications, 184(27), 35-40.

[8] Gandhi, P. N., Jangale, A., Mhaske, G. S., & Kadlag, A. (2020). *Employee Attrition Prediction Using Machine Learning.* Journal of Emerging Technologies and Innovative Research, 7(9), 1-9.

[9] Ahmed, M. N., & Moawad, I. F. (2021). *Sentiment Analysis of Employee Feedback for Enhanced Retention Strategies.* Journal of Business Analytics, 15(3), 251-266.

[10] Raj, S., Kumar, P. (2019). *Predicting Employee Turnover Using Machine Learning Algorithms: A Study on IT Sector.* International Journal of Advanced Research in Computer Science and Software Engineering, 9(6), 123-129.

[11] Smith, J., & Brown, L. (2018). *Leveraging Sentiment Analysis for Employee Engagement.* International Journal of Human Resource Studies, 8(4), 56-72.

[12] Lee, H., & Kim, S. (2020). *AI-Based Sentiment Analysis for Organizational Health Assessment.* Journal of Artificial Intelligence Research, 64, 1125-1140.

[13] Kumar, V., & Bansal, R. (2023). *Real-Time Sentiment Analysis in HR Analytics.* International Journal of Recent Technology and Engineering, 11(2), 452-459.

[14] Patel, R., & Sharma, D. (2022). *Enhancing Employee Retention Through AI and Predictive Analytics.* International Journal of Human Resource Management, 24(1), 45-58.

[15] Brown, K. and Wilson, J. (2023). *Integrating machine learning and sentiment analysis to enhance employee retention.* International Journal of Applied Engineering Research, 18(5), 345-353