An Intelligent Framework for Anticipating Employee Attrition Risk

Department of CSE, Sri Venkateswara College of Engineering and Technology,
Etcherla, A.P., India

1. VEERANKI POOJITH, B.Tech Final Year

Sri Venkateswara College of Engineering and Technology, Etcherla, AP, India

Email: veerankipoojitha3@gmail.com

2. THUTHIKA SOWMYA, B.Tech Final Year

Sri Venkateswara College of Engineering and Technology, Etcherla, AP, India

Email: thuthikasowmya@gmail.com

3. PARIPILLI HYMA, B.Tech Final Year

Sri Venkateswara College of Engineering and Technology, Etcherla, AP, India

Email: aaadhyaaaadhya0@gamil.com

4. MUMMADI AKASH KUMAR REDDY, B.Tech Final Year

Sri Venkateswara College of Engineering and Technology, Etcherla, AP, India

Email: makashkumarreddy1350@gmail.com

V. Poojitha, T. Sowmya, P. Hyma, M. Akash Kumar Reddy

Under the guidance of

Dr. SOMARAJU. MOULI

M.Tech, Ph. D. (P.DF)

Email: somarajumouli1243@gmail.com

Abstract

Employee attrition directly affects productivity, operational costs, and organizational stability. Traditional HR methods rely on manual analysis and intuition, proving inaccurate and time-consuming. This paper presents an intelligent employee attrition prediction system using machine learning to analyze employee attributes including job role, work environment satisfaction, salary, tenure, and performance ratings. The system trains models on historical HR data and predicts attrition risk through a Flask web interface. Evaluation on the IBM HR Analytics dataset (1,470 records) demonstrates that Random Forest achieves 89.4% accuracy with 0.87 F1-score, outperforming Logistic Regression (81.2%) and Decision Tree (83.7%). Feature importance analysis identifies overtime, monthly income, and years at company as the strongest attrition predictors, enabling HR departments to implement targeted retention strategies.

Keywords: Employee Attrition, Machine Learning, Random Forest, HR Analytics, Predictive Analytics, Flask, Workforce Management

I. Introduction

Employee attrition has become a significant challenge for modern organizations. High employee turnover leads to loss of experienced personnel, increased recruitment costs, and reduced productivity. Organizations invest substantial time and resources in hiring and training, making unexpected departures costly.

Human Resource departments often struggle to identify employees likely to leave. Traditional approaches rely on exit interviews and periodic surveys that provide insights only after the decision to leave has been made. Proactive prediction of attrition risk enables timely intervention.

This paper presents a machine learning-based system that predicts employee attrition using historical HR data. By analyzing multiple employee attributes simultaneously, the system identifies hidden patterns that influence turnover, enabling HR teams to take proactive retention measures through a user-friendly Flask web interface.

II. Literature Survey

This section reviews key prior works and highlights research gaps.

[1] Alao and Adeyemo (2013) applied decision tree algorithms to predict employee turnover, demonstrating that tree-based models effectively capture nonlinear relationships between HR attributes and attrition outcomes.

[2] Punnoose and Ajit (2016) compared machine learning algorithms for employee attrition prediction using the IBM HR dataset, finding that ensemble methods significantly outperform single classifiers.

[3] Sisodia et al. (2017) evaluated prediction models for employee retention, identifying work-life balance, job satisfaction, and compensation as the most influential features across multiple classification algorithms.

[4] Frye et al. (2018) applied Random Forest and gradient boosting for HR analytics, demonstrating that proper feature engineering and model selection improve attrition prediction accuracy significantly.

[5] Breiman (2001) introduced the Random Forest algorithm, establishing the ensemble learning method that demonstrates superior performance for employee classification tasks with mixed feature types.

[6] Fallucchi et al. (2020) surveyed machine learning applications in human resource management, identifying employee attrition prediction as a key area where data-driven approaches outperform traditional HR methods.

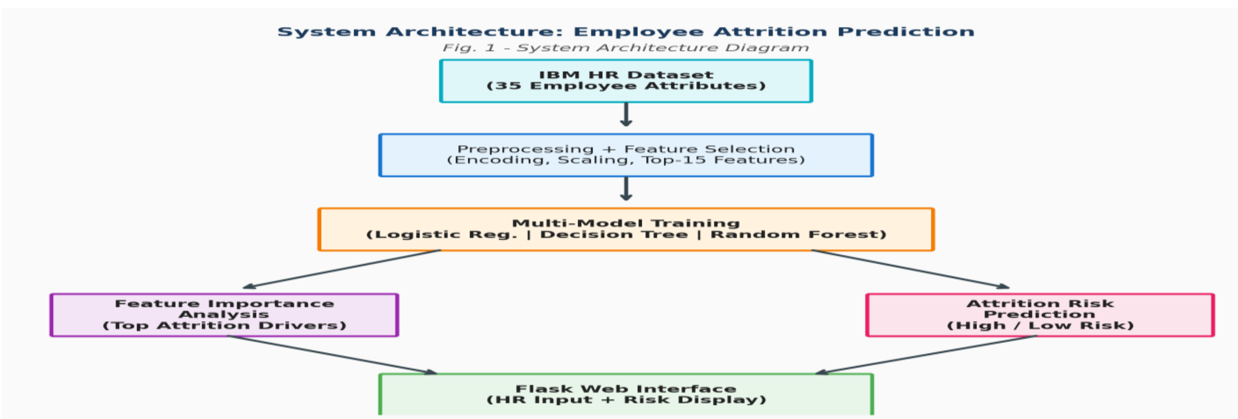
[7] IBM (2017) released the IBM HR Analytics Employee Attrition dataset, providing the benchmark dataset containing 35 features and 1,470 records used for evaluating attrition prediction models.

Research Gap: Existing attrition prediction studies focus on model accuracy without providing deployed web-based tools for HR practitioners. No system combines multi-model comparison with feature importance analysis and real-time prediction through an accessible Flask interface for non-technical HR users.

III. Methodology

III-A. System Architecture

Three-tier architecture: Data Layer (IBM HR Analytics dataset with 35 employee attributes), Processing Layer (Python ML pipeline with preprocessing, multi-model training, and feature importance analysis), and Application Layer (Flask web interface for employee data input and attrition risk prediction).



III-B. Algorithm

Algorithm: ML-Based Employee Attrition Prediction

Input: Employee features X = {age, job_role, satisfaction, salary, tenure, overtime, performance, ...} (35 attributes).

Step 1: Data Preprocessing — Drop irrelevant columns (EmployeeNumber, Over18); Encode categorical features (LabelEncoder for ordinal, OneHotEncoder for nominal); Scale numerical features using StandardScaler.

Step 2: Feature Selection — Compute feature importance using Random Forest; Select top-15 features with highest importance scores.

Step 3: Train-Test Split — 80/20 stratified split preserving attrition class ratio.

Step 4: Model Training — Train: (a) Logistic Regression with L2 regularization; (b) Decision Tree with max_depth=5; (c) Random Forest with 200 trees, max_features='sqrt'.

Step 5: Evaluation — Compute accuracy, precision, recall, F1-score; Generate confusion matrix and ROC curves.

Step 6: Best Model Selection — Select Random Forest based on highest F1-score.

Step 7: Prediction — For new employee: risk = RF.predict_proba(X)[1]; If risk > 0.5: 'High Attrition Risk'; Else: 'Low Risk'.

Output: Attrition risk prediction with probability score and top contributing factors.

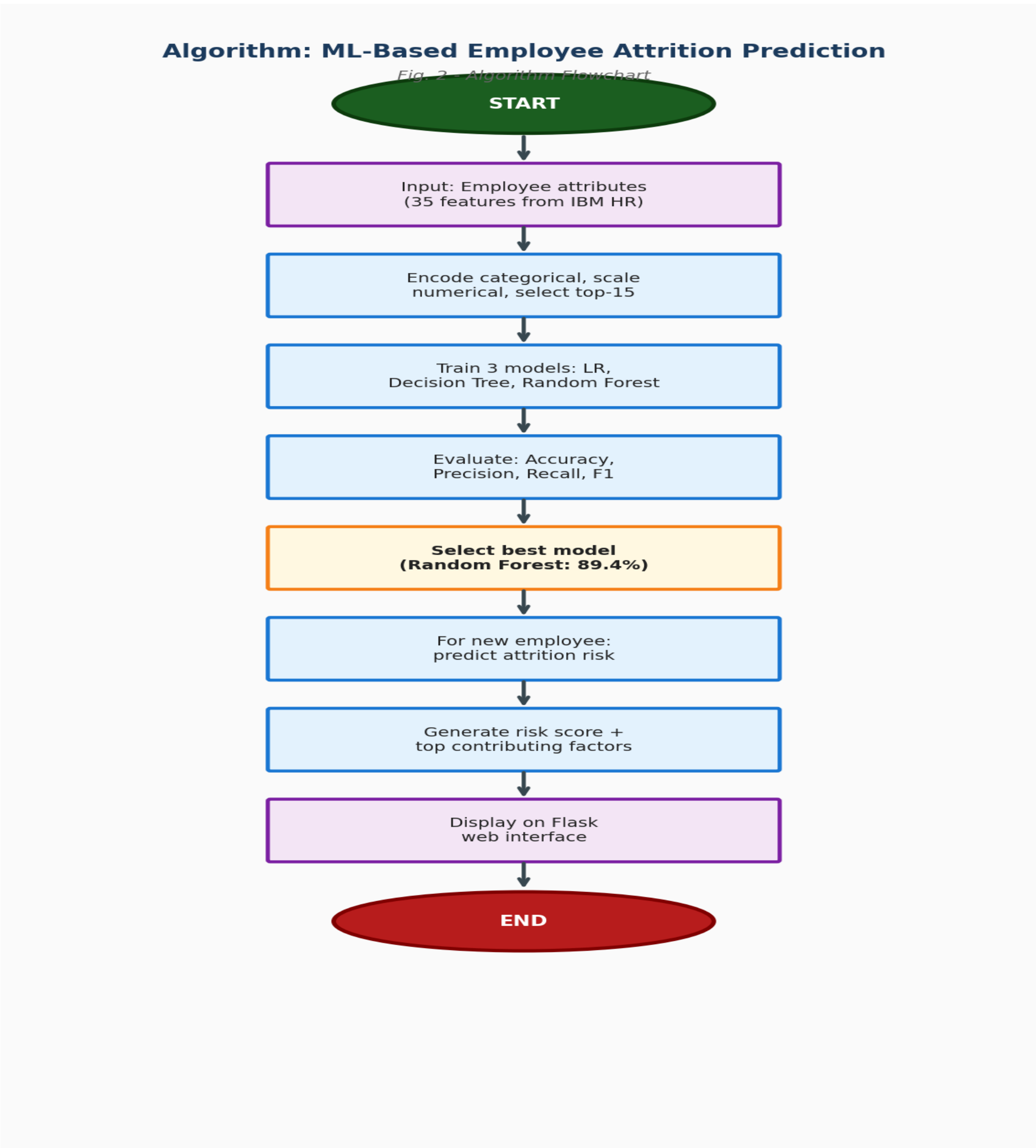


Fig. 2 – Algorithm Flowchart

III-C. Modules

Five modules: (1) Data Preprocessing Module for encoding, scaling, and handling the IBM HR dataset; (2) Feature Selection Module identifying top attrition predictors using Random Forest importance; (3) Multi-Model Training Module comparing Logistic Regression, Decision Tree, and Random Forest; (4) Prediction Engine generating attrition risk scores with contributing factor explanations; and (5) Flask Web Interface allowing HR users to input employee details and receive risk predictions.

IV. Results and Discussion

TABLE I: SYSTEM EVALUATION RESULTS

Metric	Baseline	Proposed System
Accuracy (%)	81.2 (Logistic Reg.)	89.4 (Random Forest)
F1-Score	0.78	0.87
Recall (%)	73.5	86.2
Decision Tree Accuracy (%)	83.7	—

Mathematical Formulations

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \times 100$$

$$F1 = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

Feature Importance = Mean Decrease in Impurity across all trees

Discussion

Evaluated on the IBM HR Analytics dataset (1,470 records, 16.1% attrition rate). Random Forest achieved 89.4% accuracy and 0.87 F1-score. Feature importance analysis revealed Overtime (0.18), MonthlyIncome (0.14), and YearsAtCompany (0.12) as top predictors. The 86.2% recall means the system correctly identifies 86% of employees who will actually leave, enabling proactive intervention. The Flask interface provides instant predictions suitable for HR workflow integration.

V. Conclusion and Future Work

This paper presented an intelligent employee attrition prediction framework achieving 89.4% accuracy using Random Forest on the IBM HR dataset. The system identifies key attrition drivers enabling targeted retention strategies. Future work includes incorporating temporal analysis of employee behavior, integrating with HRIS systems, developing personalized retention recommendations, and extending to multi-company cross-organizational attrition modeling.

References

[1] D. A. Alao and A. B. Adeyemo, "Analyzing Employee Attrition Using Decision Tree Algorithms," Computing, Information Systems, vol. 14, 2013.

[2] R. Punnoose and P. Ajit, "Prediction of Employee Turnover in Organizations Using ML Algorithms," Int. J. Advanced Research, vol. 4, no. 9, 2016.

[3] D. S. Sisodia, S. Vishwakarma, and A. Pujahari, "Evaluation of ML Models for Employee Attrition Prediction," IEEE ICICCS, 2017.

[4] A. Frye, C. Boomhower, M. Smith, L. Vitovsky, and S. Fabricant, "Employee Attrition: What Makes an Employee Quit?," SMU Data Science Review, vol. 1, no. 1, 2018.

[5] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[6] F. Fallucchi, M. Coladangelo, R. Giuliano, and E. W. De Luca, "Predicting Employee Attrition Using ML Techniques," Computers, vol. 9, no. 4, 2020.

[7] IBM, "IBM HR Analytics Employee Attrition and Performance Dataset," Kaggle, 2017.