ENHANCING EMPLOYEE RETENTION: AN IMPROVED HYBRID MODEL FOR PREDICTING EMPLOYEE ATTRITION

Madhubhashini J M M

239333B

MSc in Computer Science

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

FACULTY OF ENGINEERING

University of Moratuwa

Sri Lanka

June 2025

ENHANCING EMPLOYEE RETENTION: AN IMPROVED HYBRID MODEL FOR PREDICTING EMPLOYEE ATTRITION

Madhubhashini J M M

239333B

Thesis/Dissertation submitted in partial fulfillment of the requirements for the degree

MSc in Computer Science

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

FACULTY OF ENGINEERING

University of Moratuwa

Sri Lanka

June 2025

DECLARATION

I confirm that this thesis/dissertation is entirely my own work. It does not include, without proper acknowledgment, any content that has previously been submitted for any degree or diploma at any other university or higher education institution. To the best of my knowledge, it contains no material previously published or written by another individual, except where references are clearly provided. I reserve the right to reuse this work in whole or in part in future publications such as books or journal articles.

Date:27-June-2025

The above candidate has carried out research for the PhD/MPhil/Masters thesis/dissertation under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Prof. Chathura R. De Silva

Signature of the Supervisor: Date:27-June-2025

DEDICATION

I would like to dedicate the research thesis,
"Enhancing Employee Retention: Hybrid Model for Predicting Employee Attrition"
To,
The Dean of Faculty of Engineering, Prof. Manatunge J.M.A
The Head of Department of Computer Science and Engineering, Dr. Uthayasanker Thayasivam
Supervisor of the Project, Prof. Chathura R. De Silva,
all the academic and non-academic staff in the Faculty of Engineering who inspired
and motivated us in various ways.

ACKNOWLEDGEMENT

Here I owe my deep gratitude to the below mentioned people who have supported and
guided me in different
ways to meet success of the project.
My supervisor Prof. Chathura R. De Silva, for the constant encouragement and guidance
and enormous support,
that he has always given to me towards fulfilling the aim of the project.
All the academic staff at the Faculty of Engineering of University of Moratuwa, who
have support, encouragement and guidance throughout my postgraduate years.
All the non-academic staff of Faculty of Engineering of University of Moratuwa, members of my family, my batch mates and friends who have given support to me in numerous ways.

ABSTRACT

Employee attrition poses a significant challenge to organizational stability, directly impacting productivity, operational continuity, and long-term strategic goals. While traditional methods of attrition management rely on retrospective analyses, machine learning (ML) offers a proactive approach to predict and mitigate workforce turnover. This study addresses the critical need for robust predictive frameworks by evaluating eight ML classifiers; Support Vector Machines with Radial Basis Function kernel, Fisher's Linear Discriminant, Logistic Regression, Multi-Layer Perceptron, Random Forests, Naive Bayes, Adaboost and XGBoost with IBM HR Analytics Employee Attrition & Performance dataset to forecast employee attrition. A novel hybrid stacking ensemble model is proposed to enhance prediction accuracy by integrating the strengths of individual classifiers. The research investigates the impact of dataset balancing and feature optimization on model performance, emphasizing the role of data preprocessing in improving predictive reliability. Models underwent hyperparameter tuning and stratified 5-fold cross-validation, with the hybrid ensemble (Logistic Regression meta-learner) achieving superior metrics: 95.81% accuracy, 96.69% Precision, 95.77% F1-score, and 99.23% ROC-AUC. Individual models exhibited strong performance, notably XGBoost and LR (93.38% accuracy), while dataset balancing improved minority-class recall by 18–22% and feature selection reduced redundancy by 30%. Key findings reveal that dataset balancing and feature engineering significantly enhance model precision, particularly for minority class identification. The stacking model's superior performance underscores the value of meta-learning in synthesizing heterogeneous classifier outputs. Practical implications include scalable frameworks for HR analytics, enabling early risk identification and personalized retention strategies. The stacking model's adaptability allows integration with HR platforms, transforming reactive practices into proactive decision-making. Methodologically, the study advances predictive human capital management by demonstrating the synergy of hybrid ML approach and interpretable outputs. By bridging technical innovation with organizational psychology, this work offers a replicable blueprint for sustainable workforce management, emphasizing real-time analytics and employee-centric interventions.

Keywords: Machine Learning (ML) Support Vector Machines (SVM). Radial Basis Function (RBF) kernel, Fisher's Linear Discriminant (FLD), Logistic Regression (LR), Multi-Layer Perceptron (MLP), Random Forests (RF), Naive Bayes (NB)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES