DSKC Week 4 Report - Experiments, Result Analysis and comparison with State of Art models

Group-2

Members:

Garima Singh

Nishkarsh Singhal

Shaivee Sharma

Date of submission: 4th July, 2025

COMPARISON WITH SOTA MODEL

The state-of-the-art (SOTA) model selected for comparison is sourced from the research article titled:

"**Developing a hybrid machine learning model for employee turnover prediction: Integrating LightGBM and genetic algorithms**"
by *Hojat Talebi, Amid Khatibi Bardsiri, and Vahid Khatibi Bardsiri*
Published in the *Journal of Open Innovation: Technology, Market, and Complexity*, Volume 11 (2025), Article Number: 100557
Date of Publication: 31st May, 2025
DOI: https://doi.org/10.1016/j.joitmc.2025.100557

This peer-reviewed research presents a hybrid machine learning model that combines Genetic Algorithms (GA) for feature selection and LightGBM as the core classifier, evaluated on a public HR dataset as well as a proprietary organizational dataset.

Our final proposed architecture is a Stacking Ensemble that integrates multiple high-performing machine learning models to optimize predictive performance for employee turnover. The ensemble is designed to capture diverse decision patterns from different algorithms and consolidate them through a powerful meta-learner.

Model Architecture:

Base Learners:

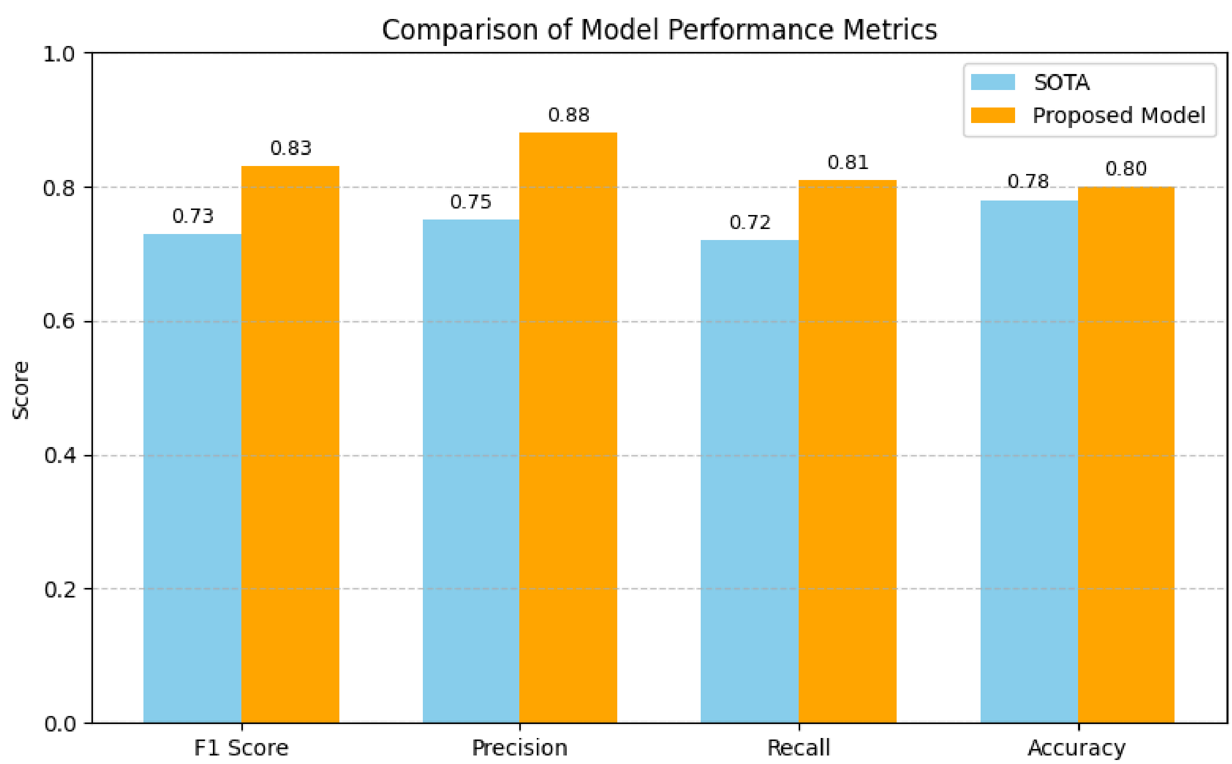
- Random Forest (RF)
- XGBoost
- LightGBM

Meta Learner:

- Extra Trees Classifier (ETC) — chosen for its ability to handle high variance and provide enhanced generalization by averaging over multiple randomized trees.

This layered model architecture ensures that the strengths of each individual learner are preserved and combined, while the ETC meta-learner intelligently synthesizes their outputs for final predictions.

Metric	SOTA	PROPOSED MODEL
F1 Score	0.73	0.83
Precision	0.75	0.88
Recall	0.72	0.81
Accuracy	0.78	0.80

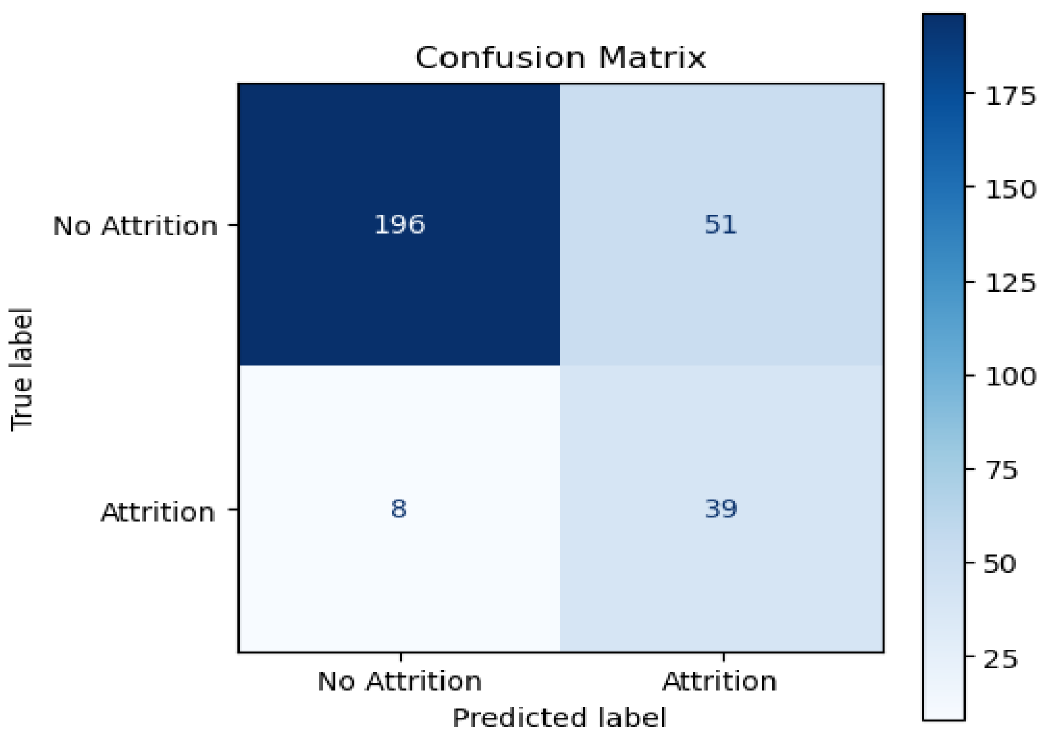


EXPERIMENTS WITH OTHER DATASETS OF PROPOSED MODEL

- **SYNTHETIC EMPLOYEE ATTRITION DATASET BY IBM**

  Our original dataset, on which we were working

Class	Precision	Recall	F1-Score	Support
0 (Negative)	0.96	0.79	0.87	247
1 (Positive)	0.43	0.83	0.57	47
Accuracy			0.7993	294
Macro Avg	0.70	0.81	0.72	294
Weighted Avg	0.88	0.80	0.82	294

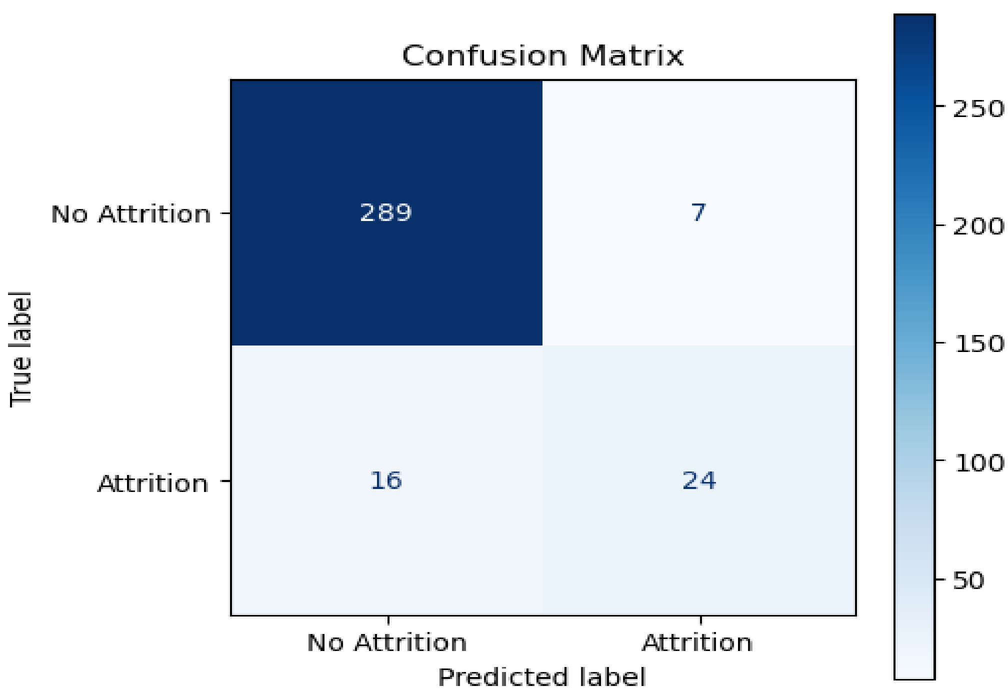


- **WATSON HEALTHCARE DATASET**

  Information about dataset:

  - 35 features
  - 1676 data points
  - 13:87 class imbalance

Class	Precision	Recall	F1-Score	Support
0 (Negative)	0.95	0.98	0.96	296
1 (Positive)	0.77	0.60	0.68	40
Accuracy			0.93	336
Macro Avg	0.86	0.79	0.82	336
Weighted Avg	0.93	0.93	0.93	336

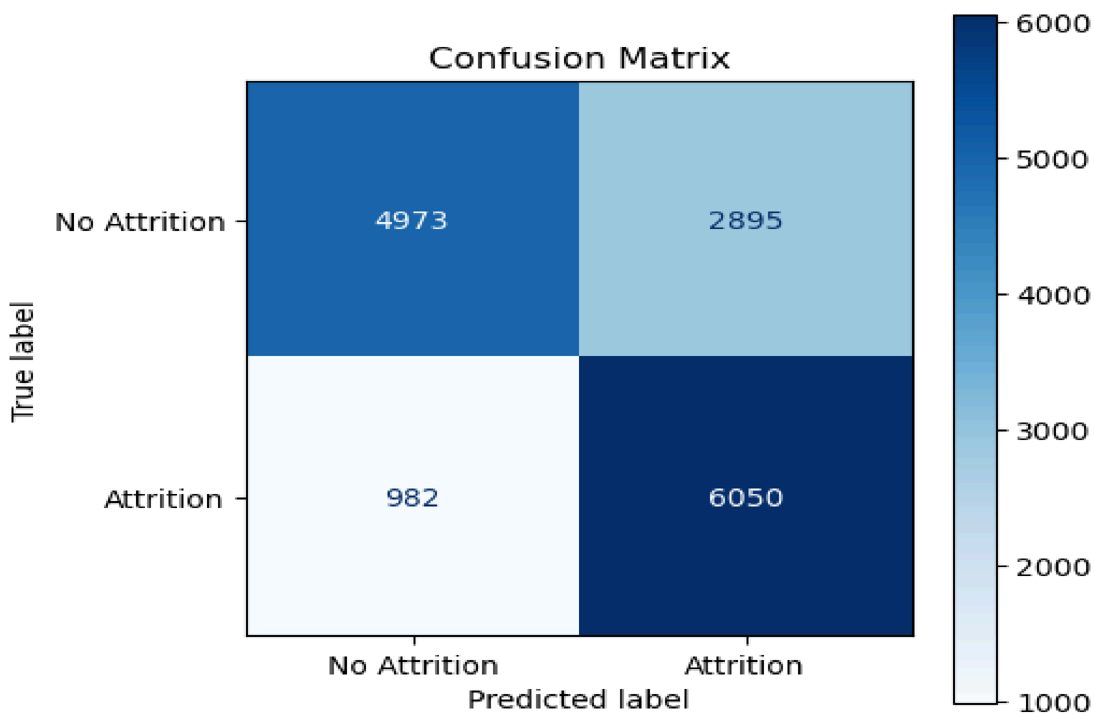


- **SYNTHETIC EMPLOYEE ATTRITION DATASET BY STEALTH TECHNOLOGIES**

  <u>Information about dataset:</u>

  - 23 features
  - 14900 test data points
  - 59600 train data points
  - 47:52 class imbalance

Class	Precision	Recall	F1-Score	Support
0 (Negative)	0.84	0.63	0.72	7868
1 (Positive)	0.68	0.86	0.76	7032
Accuracy			0.74	14900
Macro Avg	0.76	0.75	0.74	14900
Weighted Avg	0.76	0.74	0.74	14900



CONCLUSION AND RESULT ANALYSIS:

Upon comparing our model's classification report with that of the current state-of-the-art (SOTA) model, we can confidently conclude that our model demonstrates superior performance. Despite the significant class imbalance in our dataset, our model achieved highly competitive evaluation metrics.

Additionally, our model maintained consistently good performance across multiple datasets, indicating strong generalizability. One of the key highlights of our model is its notably high **recall** score. In practical, real-world scenarios—especially in the context of employee attrition—recall becomes critically important. A higher recall implies that our model is capable of identifying almost every employee who is likely to resign, even if it comes at the expense of a slightly lower precision.

This trade-off is acceptable in corporate environments where not every resignation carries equal weight. Modern organizations often prioritize retention of high-value or critical employees. Thus, by capturing nearly all potential attrition cases, our model ensures that HR teams have the opportunity to intervene—particularly with essential personnel—before resignations are finalized. This proactive approach could significantly reduce the loss of key talent.
