Application of Artificial Intelligence for Applicant Analysis

Author : Sitthisak Seemuen **Advisors :** Acting Capt. Dr. Chalermrat Nontapa and Assistant Professor Dr. Parichart Pattarapanitchai
Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai, 50200, Thailand
E-mail : Sitthisak_see@cmu.ac.th

Abstract

This study aims to develop and compare the performance of classification models for job applicant grouping and to identify the most suitable model for this task. The analysis is conducted using a dataset comprising 14,775 employees from a private company. The classification models evaluated in this study include Random Forest, XGBoost, LightGBM, and CatBoost. To address the issue of class imbalance in the dataset, three resampling techniques—over-sampling, under-sampling, and the Synthetic Minority Over-sampling Technique (SMOTE)—are applied during model development. Model performance is evaluated using a confusion matrix, with accuracy, precision, recall, and F1-score as the primary evaluation metrics. The results indicate that LightGBM outperforms the other models, achieving an accuracy of 0.9082, a precision of 0.9095, a recall of 0.9082, and an F1-score of 0.9078.

Introduction

Traditional hiring relies on subjective judgment, often leading to bias and inconsistency. In a competitive market, these inefficiencies result in poor hiring decisions that undermine organizational performance and morale.



Modern HR is shifting toward AI and Machine Learning to drive data-driven decision-making. By leveraging these technologies, organizations can eliminate bias, increase hiring accuracy, and ensure transparency. This systematic approach not only secures top talent but also fosters long-term competitive advantages.

Factors

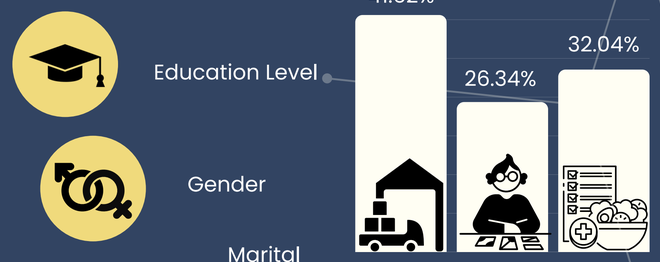
This study utilizes secondary data from 14,775 current and former employees. After rigorous data cleaning, a final dataset of 14,704 records remained for analysis.

Title / Prefix

Work Province

Religion

Employment Type



Education Level

Residential District

Gender

Work District

Marital Status

Work Shift

Commuting Distance /Distance to Work

Starting Age

Residential Province

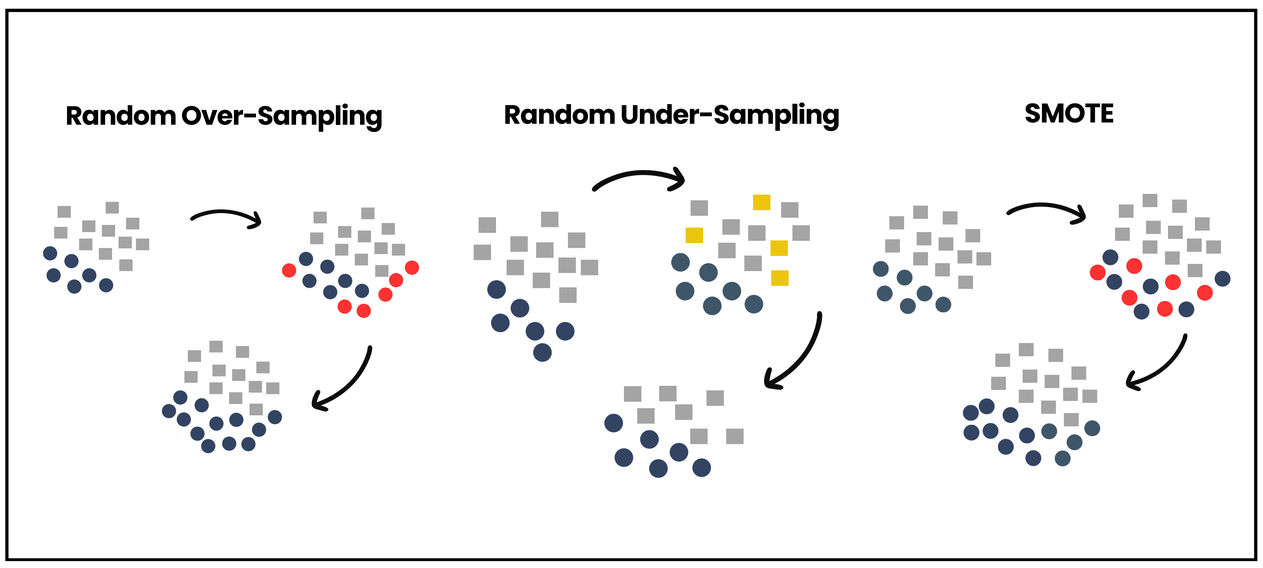
Objectives

1. To develop and compare the performance of classification models for job applicants at a specific company, utilizing Random Forest, XGBoost, LightGBM, and CatBoost algorithms.
2. To identify the most suitable model for classifying job applicants by evaluating the performance of Random Forest, XGBoost, LightGBM, and CatBoost.

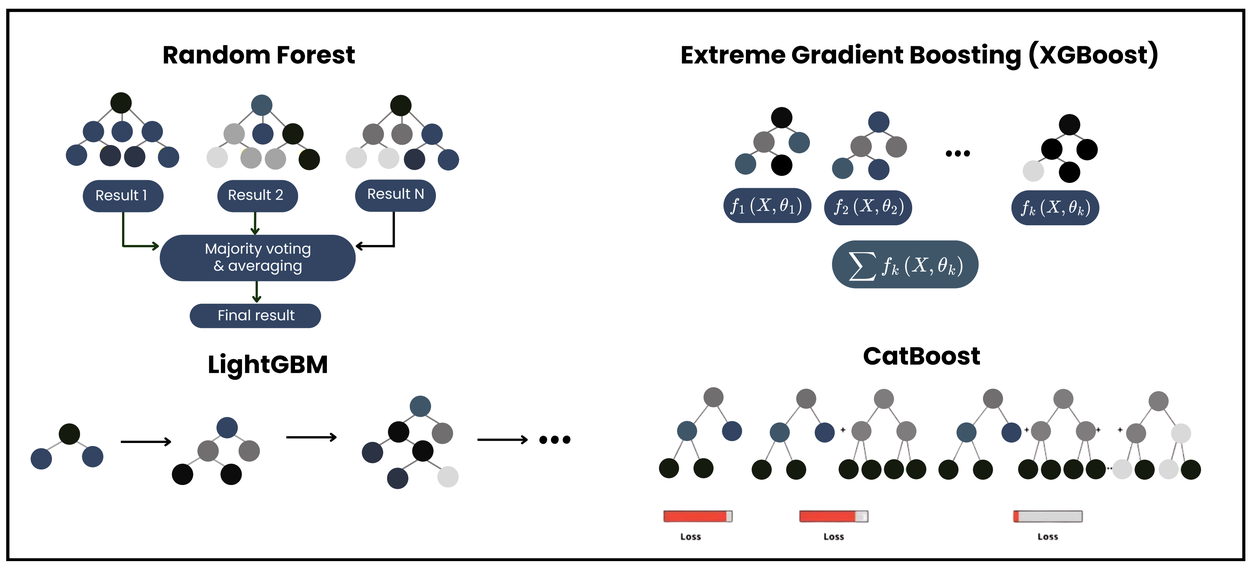
Methodology

Research and Collect Data → Data Preprocessing → Models → Imbalance Data → Summary

Imbalanced Data



Models



Evaluation Criteria

Confusion Matrix

		Actual: Positive	Actual: Negative
Predicted	Positive	TP: True Positives	FP: False Positives
Predicted	Negative	FN: False Negative	TN: True Negative

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

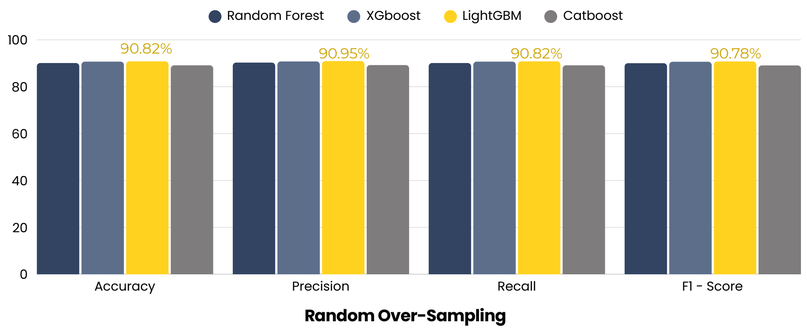
$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{F1 - Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Conclusion

The study found that LightGBM with Random Over-Sampling is the optimal model, achieving 90.82% in Accuracy and Recall, 90.95% in Precision, and 90.78% in F1-Score.



Random Over-Sampling

Accuracy	90.82%
Precision	90.95%
Recall	90.82%
F1 - Score	90.78%

Residential Province

Work District

Work Shift

Starting Age

Work Province

Discussion and Suggestions

LightGBM is the most effective model for applicant classification in this organization. Supporting the work of Peng Li (2024), it optimizes personnel matching and offers a theoretical and practical framework for human resources.

- Model Diversification: Experiment with additional approaches like Deep Learning or Hybrid Models to benchmark performance and enhance classification accuracy.
- Variable Expansion: Incorporate more detailed data (e.g., work experience, specialized skills, salary history, and overtime availability) to ensure the model is comprehensive and precise.

Benefits

Organizations can utilize these findings to improve screening procedures, thereby reducing process redundancy, decision time, and hiring costs. Furthermore, this approach increases the opportunity to secure candidates whose skills and attributes precisely match the needs of the unit.

8 DECENT WORK AND ECONOMIC GROWTH

References

[1] Rallapalli, S., & Kumar, Y. M. (2024, October). Optimizing Employee Promotion Decisions: A Novel Machine Learning Framework for Predictive Analysis by using GBM CatBoost. In 2024 First International Conference on Software, Systems and Information Technology (SSITCON) (pp. 1-7). IEEE.

[2] Li, P., & Yuan, L. (2024, October). Optimization of enterprise recruitment strategy using LightGBM decision tree: Improving talent matching accuracy. In Proceeding of the 2024 5th International Conference on Computer Science and Management Technology (pp. 1022-1027).