**International Journal of Informatics and Communication Technology (IJ-ICT)**
Vol. 15, No. 3, September 2026, pp. 1078~1086
ISSN: 2252-8776, DOI: 10.11591/ijict.v15i3.pp1078-1086 

# Integrating F-filter and K-Means SMOTE to enhance LGWUM-based turnover prediction

**Risyda Miftahur Rahmah, Sigit Priyanta**
Department of Computer Sciences and Electronics, Gadjah Mada University, Yogyakarta, Indonesia

**Article Info**

***Article history:***

Received Jul 17, 2025
Revised May 20, 2026
Accepted Jul 5, 2026

***Keywords:***

Employee turnover
F filter method
K-Means SMOTE
Lai's generalized weighted
Uplift method
Machine learning

**ABSTRACT**

High employee turnover poses a significant challenge for organizations. While uplift modeling offers a prescriptive analytics approach by estimating differential treatment effects to optimize retention programs, its performance is often hindered by irrelevant features and imbalanced class distributions. To address these issues, this study proposes an employee turnover model utilizing lai's generalized weighted uplift method (LGWUM), enhanced with F-Filter feature selection and K-Means SMOTE for a refined feature space and balanced treatment representation. Evaluated across three HR datasets with four engineered uplift classes (CN, CR, TN, TR), the integrated framework significantly improves uplift performance, yielding increased Qini coefficients of 0.0755 and 0.0870 on Datasets 2 and 3, respectively. Furthermore, top-decile probability distribution analysis confirms a clearer separation between positive and negative responders, with XGBoost demonstrating the most robust and reliable uplift discrimination across the models.



***Corresponding Author:***

Sigit Priyanta
Department of Computer Sciences and Electronics, Faculty of Mathematics and Natural Sciences
Gadjah Mada University
Bulaksumur, Kec. Depok, Kabupaten Sleman, 55281, Daerah Istimewa Yogyakarta
Email: seagatejogja@ugm.ac.id

## 1. INTRODUCTION

An effective organization understands that human resources are a key driver of sustainability and business growth [1]. In practice, employee turnover, voluntary separation between employees and the organization [2], remains a common challenge and can increase recruitment costs, reduce institutional knowledge, and disrupt team dynamics. To anticipate turnover, research by [3] emphasizes the value of HR analytics for strategic HR planning. One of the most effective solutions is implementing retention programs, which help maintain competent talent, increase productivity, and reduce turnover-related costs [4].

Traditional predictive models use binary classification to identify employees likely to leave by distinguishing between those who resign and those who stay, enabling organizations to prioritize high-risk employees for retention [5]. In these models, the target variable is binary $y \in \{0, 1\}$. However, traditional models cannot capture the impact of retention interventions, while uplift modeling estimates the increased probability of employees staying due to treatment [6].

In the context of employee turnover, uplift modeling compares randomly assigned treatment and control groups to evaluate retention effectiveness. Instead of just predicting who will leave, this approach pinpoint the "persuadable" employees who are most likely to respond positively to interventions [7]. An uplift model is a predictive model designed to estimate the effect of taking a particular action on a specific outcome, also known as the estimation of conditional average treatment effects (CATE) [8].

A previous study by [9], [10] compared traditional predictive and uplift models for employee turnover using lai's generalized weighted uplift method (LGWUM). The traditional model uses a binary churn target, while the uplift model applies a four-class treatment-control target, enabling better identification of Persuadable employees and more efficient retention strategies.

A major challenge in uplift modeling is high feature complexity, making feature selection essential to reduce overfitting, improve efficiency, and enhance interpretability [11]. However, traditional feature selection methods are less suitable for uplift modeling because they focus on outcome prediction rather than treatment effect differences between treated and untreated groups [12]. Zhao *et al.* [13] introduced feature selection methods for uplift modeling, including F-filter, LR-filter, and embedded approaches, have been introduced to select the most relevant features, improving efficiency and overall model performance.

The second challenge is dataset imbalance, which hinders learning because most supervised algorithms assume balanced classes. Although several solutions exist, synthetic oversampling is often preferred as it adjusts the data rather than the classifier. SMOTE [14], is a common oversampling method that generates synthetic minority samples but is sensitive to noise and random sampling [15]. To overcome these weaknesses, [16] introduced K-means SMOTE, which improves the original SMOTE through a three-stage process: clustering, filtering, and oversampling.

Numerous studies show that combining feature selection and oversampling effectively addresses class imbalance and improves predictive performance. Huang *et al.* [17] demonstrated that the integration of information gain-based feature selection with SMOTE outperformed the individual application of either technique, particularly when feature selection preceded oversampling.

As this study uses uplift modeling, feature selection targets heterogeneous treatment effects. Therefore, the F-Filter is combined with K-Means SMOTE to address imbalance across treatment and control groups. To our knowledge, this combination has not been explored in uplift-based employee turnover prediction, representing a novel contribution.

This study makes three key contributions: (1) applying the uplift-specific F-Filter to identify features driving heterogeneous treatment effects in turnover; (2) integrating K-Means SMOTE to mitigate treatment-control imbalance and stabilize uplift; and (3) evaluating the approach across multiple machine learning models for robustness. Collectively, these contributions advance prescriptive analytics for employee retention, improving uplift performance and enabling efficient, personalized intervention allocation.

## 2. THE PROPOSED METHOD

There are eight uplift modeling strategies. In this study, we adopt LGWUM that proposed by [18], to ensure balanced response measurement, uplift estimation weights probabilities based on treatment and control proportions. Here, the treatment is defined as the employee retention program, dividing subjects into the Treatment Group (T) and the Control Group (C). Employee responses are classified across two dimensions, response outcome (yes/no) and treatment assignment (yes/no), resulting in four distinct outcome groups visualized in Figure 1.

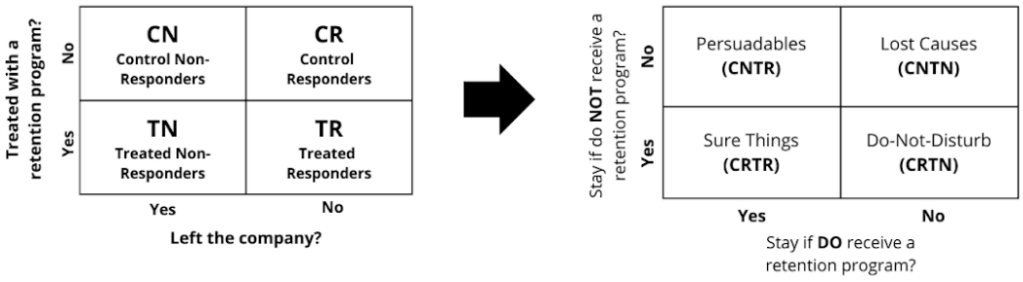



Figure 1. Mapping employee responses and retention program assignment to the four uplift classes

Employee Group Classification Based on Treatment Response [9]:

a) Control Non-Responders (CN): Employees who did not receive treatment and left the company.
b) Control Responders (CR): Employees who stayed without receiving treatment.
c) Treated Non-Responders (TN): Employees who received treatment but still left.
d) Treated Responders (TR): Employees who stayed after receiving treatment, the objective is to detect persuadable employees within this group who truly benefited from treatment.

After using a machine learning algorithm to predict the four target classes, four probability values are produced. Let P represent these probabilities; the uplift score is then computed using the LGWUM method as follows [9] (1):

$$Uplift\ Score = \frac{P(CN \mid x)}{P(C)}) + \frac{P(TR \mid x)}{P(T)}) - \frac{P(CR \mid x)}{P(C)}) - \frac{P(TN \mid x)}{P(T)} \quad (1)$$

The uplift score is high when P(CN|x) and P(TR|x) are high, and P(CR|x) and P(TN|x) are low, indicating that the employee likely belongs to the Persuadables group and should be prioritized for the retention program. Employees are categorized into four uplift classes:

a) Persuadables: Stay if treated; leave if not (primary target).
b) Sure Things: Stay regardless of treatment.
c) Lost Causes: Leave regardless of treatment.
d) Do-Not-Disturb: Driven away (leave) if treated.

### 2.1. F-Filter: method for selecting features in uplift modeling

Feature selection plays a more critical role in uplift modeling than in standard predictive modeling [12]. Zhao *et al.* [13] introduced the F-Filter and LR-Filter, which evaluate feature relevance using p-values and statistical significance. In the F-Filter, feature importance is quantified using the F-statistic as defined in (2).

$$F = \frac{(RSS - RSS\prime) / R}{RSS\prime / (N - R - 2)} \quad (2)$$

In this formulation, $N$ is the total number of observations, $RSS'$ is the Residual Sum of Squares of the model with estimated coefficients, and $RSS$ represents the baseline model. $I$ is the treatment indicator (1 = treatment, 0 = control), $R$ is the polynomial degree, $x_j^r$ denotes the $x_j$ feature raised to the $r$-th power, and $\varepsilon$ is the error term. The parameters $\alpha$, $\beta$, and $\theta$ are model coefficients, with $\theta$ indicating the magnitude of the heterogeneous treatment effect of feature $x_j$.

### 2.2. K-Means SMOTE

K-Means SMOTE, introduced by [16], involves clustering the data using K-means, selecting minority-dense clusters for oversampling, and applying SMOTE within those clusters. Synthetic samples are generated proportionally, prioritizing clusters with sparser minority distributions to improve class balance. K-Means is an iterative clustering algorithm that assigns points to the nearest centroid and updates centroids as the mean of assigned points until convergence to a local optimum [19]. Hairani *et al.* [20] have shown that K-Means SMOTE enhances classification performance and minority sample quality while preserving data structure, albeit with increased computational cost.

This study combines F-Filter feature selection and K-Means SMOTE to improve uplift modeling. After splitting the data, F-Filter selects features related to treatment-effect heterogeneity, and K-Means SMOTE is applied separately to treatment and control groups to handle imbalance. The balanced data are then used to train models, and LGWUM computes uplift scores to estimate individual treatment effects.

## 3. RESEARCH METHOD

This study involves data cleaning, feature engineering to create additional features based on treatment and control groups, as well as uplift response features. The research methodology is illustrated in Figure 2.

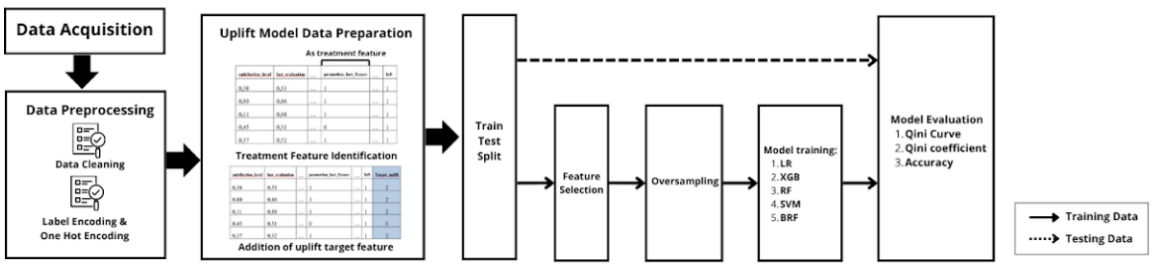



Figure 2. Flowchart and proposed method

### 3.1. Data acquisition

Dataset 1 is a synthetic human resources dataset with 10 features and a binary target variable (*Left*), consisting of 3,571 employees who left and 11,428 who did not, for a total of 14,999 records. In [21], Dataset 2 is a synthetic dataset from IBM Watson Analytics with 35 features and a binary target (*Attrition*), containing 237 employees who left and 1,233 who did not, totaling 1,470 records [22], and Dataset 3 is a real-world dataset with 16 features and a binary target (*Event*), comprising 426 employees who left and 389 who did not, resulting in 815 records [23].

### 3.2. Data preprocessing

Data preprocessing was performed by handling missing values and transforming categorical data into numerical form using label encoding and one-hot encoding, increasing the number of features from 10 to 19 in Dataset 1, 31 to 50 in Dataset 2, and 16 to 52 in Dataset 3.

### 3.3. Uplift model data preparation

Since the datasets lack an inherent treatment-control feature, the treatment variable was determined based on three criteria from [9]: actionability, correlation with the target, and control group availability. Based on these criteria, the selected treatment variables for each dataset are "Promotion" for Dataset 1 (comprising 319 treatment and 14,680 control samples; treatment correlation: 0.062), "Overtime" for Dataset 2 (1,054 treatment and 416 control samples; treatment correlation: 0.246), and "Coaching" for Dataset 3 (683 treatment and 132 control samples; treatment correlation: 0.064). Prior studies show these interventions can reduce turnover and improve employee retention. Promotion increases morale and performance, reducing turnover [24], managing overtime lowers job stress and turnover intention [25], and coaching improves engagement and commitment, strengthening retention [26]. The model uses four encoded target classes: CN, CR, TN, and T, which are subsequently encoded as 0, 1, 2, and 3, respectively. After feature engineering, the dataset was split into training and testing sets using a 70:30 ratio.

### 3.4. Feature selection

Feature selection was conducted by ranking features using F-scores and p-values, then iteratively removing the lowest-ranked features in steps of five. Each subset was evaluated on the test data using XGBoost to assess changes in uplift performance via the Qini coefficient. Feature importance was also reported at each stage to identify influential predictors. Results are summarized in Table 1.

Table 1. Experiment to determine the optimal number of features in dataset 1

| Dataset | Metrics | All Features | 45 | 40 | 35 | 30 | 25 | 20 | 15 | 10 | 5 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Dataset 1 (16) | Qini Coefficient | 0,1855 | - | - | - | - | - | - | 0,1849 | 0,1260 | 0,1040 |
| | FI-Score | 0,032050 | - | - | - | - | - | - | 0,077765 | 0,723418 | 3,331304 |
| Dataset 2 (47) | Qini Coefficient | 0,0479 | 0,0485 | 0,0267 | 0.0423 | 0,0634 | 0,0295 | 0,0165 | 0,0128 | 0,0114 | -0,0084 |
| | FI-Score | 0,000083 | 0,000257 | 0,078256 | 0,237678 | 0,812167 | 1,130739 | 1,442817 | 2,647085 | 4,017086 | 8,627038 |
| Dataset 3 (49) | Qini Coefficient | 0,0608 | 0,0605 | 0,0683 | 0,0754 | 0,0333 | 0,0567 | 0,0523 | 0,0723 | 0,0429 | 0,0302 |
| | FI-Score | 0,004306 | 0,015455 | 0,071298 | 0,131229 | 0,313057 | 0,430212 | 0,892153 | 1,341702 | 1,873390 | 2,647091 |

In Dataset 1, feature selection was limited to 15 features due to its smaller size (16 features), compared with Dataset 2 (47 features) and Dataset 3 (49 features). Gradual feature reduction in steps of five shows that removing less informative features improves uplift performance up to an optimal point. Beyond this threshold, eliminating influential predictors reduces the Qini coefficient, as reflected by rising feature importance values. Overall, results indicate that each dataset has an optimal feature subset where uplift performance is maximized using the minimal set of impactful features.

### 3.5. Handling imbalanced data

Before modeling, the class imbalance was addressed using the K-Means SMOTE method [14]. To address class imbalance, K-Means SMOTE was applied across three datasets with varying strategies based on the severity of the imbalance. In Dataset 1, which suffered from severe imbalance, minority classes

were oversampled conservatively but remained stratified (increasing Class 0 from 13 to 300, Class 1 from 210 to 1,200, and Class 2 from 2,479 to 3,000) against a majority Class 3 of 7,797. Conversely, for Datasets 2 and 3 which exhibited moderate imbalance, all minority classes were oversampled to successfully match their respective majority class baselines, bringing Dataset 2 classes to an even distribution of approximately 659 instances each, and Dataset 3 classes to roughly 257 instances each.

### 3.6. Evaluation

Traditional metrics like accuracy evaluate prediction correctness at the individual level, which is unsuitable for uplift modeling because individual uplift cannot be directly observed, one person cannot be both in the treatment and control groups. Therefore, uplift must be measured at the segment level, and the Qini coefficient, introduced in [27] and adapted from the Gini coefficient, is used to evaluate uplift performance. Through equation, uplift can be normalized in (3).

$$Uplift(\alpha) = N\alpha(\frac{TR}{T} - \frac{CR}{C}) \quad (3)$$

Here, $\alpha$ represents the proportion of targets in the treatment group, and N denotes the total population. A Seaborn line plot is used to visualize the Qini curve for the uplift model, with a random model curve added for comparison [9] (4).

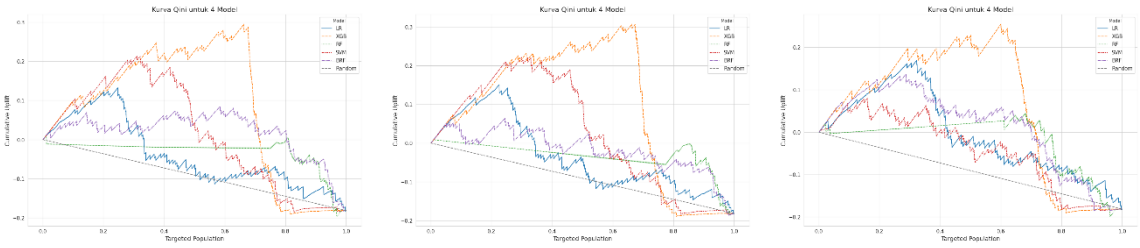
$$qini\ coefficient = \sum_{i=0}^{N-1} uplift - random\ model\ curve \quad (4)$$

To evaluate whether a model successfully targets the right employees, the model's Qini curve is compared with that of a random model. Therefore, a model is considered effective if it achieves a Qini coefficient greater than 0.05 [9].

## 4. RESULTS AND DISCUSSION

### 4.1. Qini curve and qini coefficient

The study utilizes five machine learning models (Logistic Regression, XGBoost, Random Forest, SVM, and Balanced Random Forest) to generate class probabilities, which are then converted into ranked uplift scores. Ultimately, model performance is evaluated against a random benchmark using the Qini curve and Qini coefficient, serving as the standard evaluation metrics for uplift modeling [28]. Figure 3 displays the Qini curve and Qini coefficient for Dataset 1.



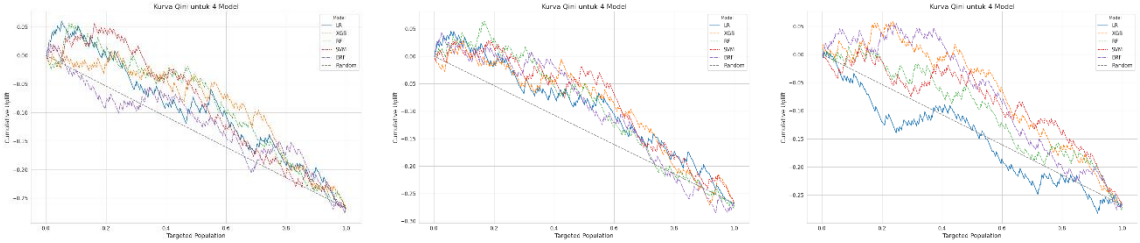

Without Feature Selection (16 features)
LR: 0.0558
XGB: 0.1855
RF: 0.0519
SVM: 0.1118
BRF: 0.0726

Feature Selection (15 features)
LR: 0,0497
XGB: 0.1686
RF: 0.0544
SVM: 0.1028
BRF: 0.1076

Feature Selection (15 features) + KMeans SMOTE
LR: 0,0845
XGB: 0,1490
RF: 0,0732
SVM: 0,0479
BRF: 0,1023

Figure 3. Qini curve and qini coefficient for dataset 1

For Dataset 1, XGBoost consistently achieves the highest Qini coefficient, demonstrating strong capability in capturing heterogeneous treatment effects. Feature selection alone slightly degrades performance, while its combination with K-Means SMOTE improves uplift results for several models, underscoring the importance of balanced treatment representation. Figure 4 shows the Qini curve and coefficient for Dataset 2.



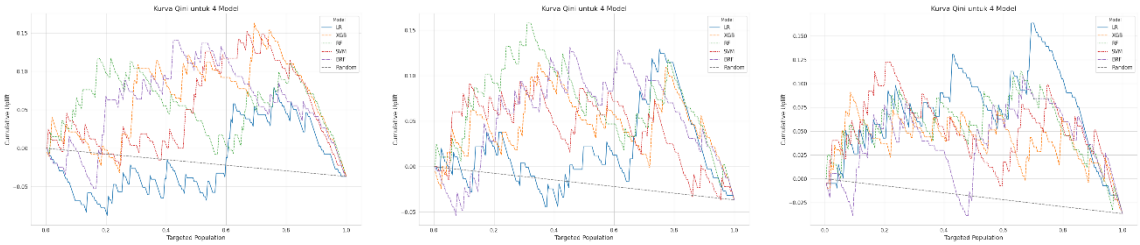
Without Feature Selection
LR: 0,0394
XGB: 0,0479
RF: 0,0494
SVM: 0,0459
BRF: 0,0144

Feature Selection (30 features)
LR: 0,0430
XGB: 0,0634
RF: 0,0457
SVM: 0,0578
BRF: 0,0304

Feature Selection (30 features) + KMeans SMOTE
LR: -0,0112
XGB: 0,0755
RF: 0,0487
SVM: 0,0474
BRF:0,0455

Figure 4. Qini curve and qini coefficient for dataset 2

For Dataset 2, uplift performance improves with feature selection, indicating that removing irrelevant features enhances treatment-effect learning. Combining feature selection with K-Means SMOTE produces the highest Qini scores, with XGBoost again performing best, demonstrating robustness to both dimensionality reduction and class imbalance. Figure 5 presents the Qini curve and Qini coefficient for Dataset 3.



Without Feature Selection
LR: 0.0015
XGB: 0.0832
RF: 0.0736
SVM: 0.0685
BRF: 0.0755

Feature Selection (35 features)
LR: 0,0353
XGB: 0.0649
RF: 0.0865
SVM: 0.0598
BRF: 0.0726

Feature Selection (35 features) + KMeans SMOTE
LR: 0,0870
XGB: 0,0577
RF: 0,0726
SVM: 0,0689
BRF: 0,0569

Figure 5. Qini curve and qini coefficient for dataset 3

For Dataset 3, feature selection improves uplift performance across most models, while adding K-Means SMOTE provides little benefit and slightly reduces Qini scores. Overall, tree-based models perform well, with uplift gains driven mainly by feature selection rather than oversampling. Overall, integrating F-Filter and K-Means SMOTE improves uplift performance by preserving treatment-response heterogeneity, especially in Dataset 2 and 3, showing that tailored preprocessing enhances model effectiveness.

### 4.2. Model performance discussion

The uplift model is primarily evaluated using the Qini coefficient, while accuracy is included as a supporting metric to assess how feature selection and K-Means SMOTE affect class-label prediction performance. Table 2 reports the best Qini coefficients obtained in this study and compares them with the findings of previous research by [9] across all three datasets.

Table 2 demonstrates that the proposed method successfully improves Qini coefficients across all datasets. Uplift-specific feature selection consistently enhances performance, while K-Means SMOTE provides additional gains in some datasets and marginal benefits in others. Overall, these results confirm that the proposed approach more effectively improves uplift modeling performance than previous research. Evaluation using accuracy, as reported in Table 3.

Table 2. Comparison of qini coefficient values with previous research

| Dataset | Previous Research [9] | | Proposed Method | | | |
|---|---|---|---|---|---|---|
| | | | Without K-Means SMOTE | | With K-Means SMOTE | |
| | Method | Qini coefficient | Method | Koefisien qini | Method | Koefisien Qini |
| Dataset 1 | XGB | 0,1240 | XGB | 0,1855 | XGB | 0.1760 |
| Dataset 2 | XGB | 0.0393 | XGB + F Filter (30 features) | 0,0634 | XGB + F Filter (30 features) | 0,0755 |
| Dataset 3 | XGB | 0.0697 | RF + F Filter (35features) | 0,0865 | LR + F Filter (35 features) | 0,0870 |

Table 3. Comparison of accuracy values with previous research

| Dataset | Previous Research [9] | | Proposed Method | | | |
|---|---|---|---|---|---|---|
| | | | Without K-Means SMOTE | | With K-Means SMOTE | |
| | Method | Accuracy | Method | Accuracy | Method | Accuracy |
| Dataset 1 | XGB | 96,37% | RF + F Filter (10 features) | 97,84% | RF + F Filter (10 features) | 97,64% |
| Dataset 2 | XGB | 62,25% | LR | 67,12% | RF + F Filter (35 features) | 65,08% |
| Dataset 3 | XGB | 51,48% | RF + F Filter (35 features) | 57,55% | BRF | 58,37% |

Table 3 shows that the proposed method improves accuracy across all datasets compared to previous research. Feature selection consistently enhances accuracy, while the benefit of K-Means SMOTE varies, providing gains mainly in highly imbalanced data.

### 4.3. Analysis of class probability distribution at top decile

The uplift model aims to rank individuals by their likelihood of benefiting from treatment, focusing on the top decile to guide cost-effective targeting [29]. This study analyzed class probability distributions within the top 10% of uplift scores for the best-performing models from Table 2: Dataset 1 (no feature selection and no oversampling), Dataset 2 (30 features and oversampling), and Dataset 3 (35 features and oversampling). Figure 6 illustrates the distribution of class probability at top decile for Dataset 1, Dataset 2, and Dataset 3.

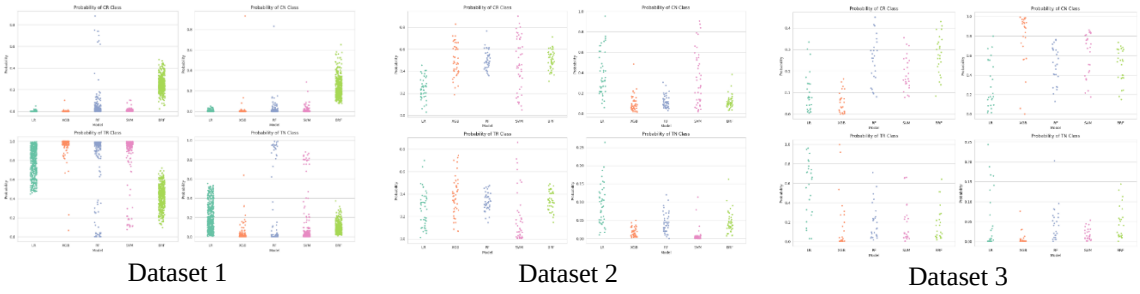



Figure 6. Distribution of class probability at top decile for Dataset 1, Dataset 1, and Dataset 3

Top-decile probability plots confirm that XGBoost is the most reliable model for treatment targeting compared to LR, RF, SVM, and BRF. Across all datasets, XGBoost consistently excels at identifying and prioritizing the "persuadable" TR segment while successfully avoiding the CR and TN groups to mitigate negative treatment risks, making it the superior choice for distinguishing who will truly benefit from interventions. Its regularized gradient boosting framework iteratively corrects past errors, improving separation between TR and CN while reducing misclassification of CR and TN [30]. In contrast, LR assumes linear boundaries, RF does not learn iteratively from prior errors [31], SVM focuses on margin separation rather than treatment effects, and BRF emphasizes class balancing rather than uplift optimization leading to weaker class discrimination.

## 5. CONCLUSION

Our findings suggest that combining F-Filter feature selection and K-Means SMOTE improves uplift modeling performance. The higher Qini coefficient indicates that reducing irrelevant features and balancing treatment groups enables the model to better capture treatment-response heterogeneity. We found that XGBoost remains the strongest algorithm due to its consistent ability to separate positive and negative

responders in the top decile, making it the most reliable choice for the LGWUM framework. Practically, this approach helps organizations optimize resource allocation, such as in HR retention, by targeting interventions exclusively to individuals who will genuinely benefit, thereby maximizing retention results while minimizing unnecessary costs. The findings remain dataset-dependent, meaning model performance may not generalize to other domains. Oversampling with K-Means SMOTE also introduces a risk of overfitting due to synthetic patterns, and uplift results are sensitive to how treatment features are defined, which can affect the stability of heterogeneous treatment effect estimation. Future studies may explore uplift modeling in other domains and multi-treatment settings, develop methods for determining the optimal number of selected features, enable real-time deployment, and explore fairness-aware uplift modeling.

## FUNDING INFORMATION

The authors would like to express their sincere gratitude to the Department of Computer Science and Electronics, Gadjah Mada University, for the support and facilities provided during the completion of this research.

## AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

| Name of Author | C | M | So | Va | Fo | I | R | D | O | E | Vi | Su | P | Fu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Risyda Miftahur Rahmah | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | | ✓ | | ✓ | ✓ |
| Sigit Priyanta | ✓ | ✓ | | ✓ | ✓ | ✓ | | | | ✓ | | ✓ | ✓ | ✓ |

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

## CONFLICT OF INTEREST STATEMENT

The authors state that there are no conflicts of interest related to this research.

## DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

## REFERENCES

[1] S. Sella, H. Riofita, P. Studi, P. Ekonomi, U. Islam, and S. Syarif, "The role of human resource management in achieving a company's strategic goals," *Asosiasi Riset Ilmu Manajemen dan Bisnis Indonesia*, 2024.

[2] A. Cohen and R. Golan, "Predicting absenteeism and turnover intentions by past absenteeism and work attitudes: An empirical examination of female employees in long term nursing care facilities," *Career Development International*, vol. 12, no. 5, pp. 416–432, 2007, doi: 10.1108/13620430710773745.

[3] H. C. Ben-Gal, D. Avrahami, D. Pessach, G. Singer, and B. G. Irad, "A human resources analytics examination of turnover: implication for theory and practice," *81st Annual Meeting of the Academy of Management 2021: Bringing the Manager Back in Management, AoM 2021*, no. May 2024, 2021, doi: 10.5465/AMBPP.2021.158.

[4] I. G. Kang, N. Kim, W. Y. Loh, and B. A. Bichelmeyer, "A machine-learning classification tree model of perceived organizational performance in u.S. federal government health agencies," *Sustainability (Switzerland)*, vol. 13, no. 18, 2021, doi: 10.3390/su131810329.

[5] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach," *European Journal of Operational Research*, vol. 218, no. 1, pp. 211–229, 2012, doi: 10.1016/j.ejor.2011.09.031.

[6] C. Mayes and K. K. Govender, "Exploring uplift modelling in direct marketing," *International Journal of Financial, Accounting, and Management*, vol. 1, no. 2, pp. 69–79, 2019, doi: 10.35912/ijfam.v1i2.81.

[7] V. S. Y. Lo, "The true lift model: a novel data mining approach to response modeling in database marketing," *ACM SIGKDD Explorations Newsletter*, vol. 4, no. 2, pp. 78–86, 2002.

[8] R. Gubela, A. Bequé, F. Gebert, and S. Lessmann, "Conversion uplift in e-commerce : A systematic benchmark of modeling strategies Conversion uplift in e-commerce : A systematic benchmark of modeling strategies," *International Research Training Group 1792*, vol. 2018–062, 2018.

[9] D. Wijaya, J. H. DS, S. Barus, B. Pasaribu, L. I. Sirbu, and A. Dharma, "Uplift modeling VS conventional predictive model: A reliable machine learning model to solve employee turnover," *International Journal of Artificial Intelligence Research*, vol. 5, no. 1, pp. 53–64, 2021, doi: 10.29099/ijair.v4i2.169.

[10] J. Kinoto, J. L. Damanik, E. T. S. Situmorang, J. Siregar, and M. Harahap, "Employee churn prediction with uplift modeling using the logistic regression algorithm (in Indonesian)," *Jurnal Teknologi Dan Ilmu Komputer Prima (Jutikomp)*, vol. 3, no. 2, pp. 503–508, 2020, doi: 10.34012/jutikomp.v3i2.1645.
[11] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, "A review of feature selection methods on synthetic data," *Knowledge and Information Systems*, vol. 34, no. 3, pp. 483–519, 2013, doi: 10.1007/s10115-012-0487-8.
[12] N. Radcliffe and P. Surry, "Real-world uplift modelling with significance-based uplift trees," *White Paper TR-2011-1, Stochastic*, no. section 6, pp. 1–33, 2011.
[13] Z. Zhao, Y. Zhang, T. Harinen, and M. Yung, "Feature selection methods for uplift modeling and heterogeneous treatment effect," *IFIP Advances in Information and Communication Technology*, vol. 647 IFIP, pp. 217–230, 2022, doi: 10.1007/978-3-031-08337-2_19.
[14] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.
[15] R. C. Prati, G. E. A. P. A. Batista, and M. C. Monard, "Learning with class skews and small disjuncts," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3171, no. May 2014, pp. 296–306, 2004, doi: 10.1007/978-3-540-28645-5_30.
[16] F. Last, G. Douzas, and F. Bacao, "Oversampling for imbalanced learning based on K-Means and SMOTE," *arXiv preprint arXiv:1711.00837 2*, pp. 1–19, 2017, doi: 10.1016/j.ins.2018.06.056.
[17] M. Huang, C. Chiu, C. Tsai, and W. Lin, "Applied sciences on combining feature selection and over-sampling techniques for breast cancer prediction," *Applied Sciences*, 2021.
[18] K. Kane, V. S. Y. Lo, and J. Zheng, "Mining for the truly responsive customers and prospects using true-lift modeling: comparison of new and existing methods," *Journal of Marketing Analytics*, vol. 2, no. 4, pp. 218–238, 2014, doi: 10.1057/jma.2014.18.
[19] M. J, "Some methods for classification and analysis of multivariate observations," *Proc of Berkeley Symposium on Mathematical Statistics and Probability*, vol. 5, no. 1, pp. 281–297, 1965.
[20] H. Hairani, K. E. Saputro, and S. Fadli, "K-means-SMOTE for handling class imbalance in the classification of diabetes with C4.5, SVM, and naive Bayes," *Jurnal Teknologi dan Sistem Komputer*, vol. 8, no. 2, pp. 89–93, 2020, doi: 10.14710/jtsiskom.8.2.2020.89-93.
[21] G. Pujar, "HR Analytics," *Kaggle*, 2018. .
[22] P. Subhasht, "IBM HR analytics employee attrition and performance," *Kaggle*, 2017.
[23] E. Babushkin, "Employee turnover: how to predict individual risks of quitting," 2017.
[24] U. Madugu, O. Ogundeji, "Promotion and job satisfaction : a precursor of high," no. August, 2023.
[25] A. Junaidi, E. Sasono, W. Wanuri, and D. Wahyu, "Management science letters," vol. 10, pp. 3873–3878, 2020, doi: 10.5267/j.msl.2020.7.024.
[26] M. Kamunya, "Influence of coaching on employee retention in commercial banks in Kenya," *Human Resource and Leadership Journal*, vol. 5, no. 1, pp. 29–50, 2020.
[27] N. J. Radcliffe and N. J, "Using control groups to target on predicted lift : 1 introduction," *Direct Marketing Analytics Journal*, no. 2, 2007.
[28] A. Shimizu, R. Togashi, A. Lam, and N. Van Huynh, "Uplift modeling for cost effective coupon marketing in C-to-C E-Commerce," *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, vol. 2019-Novem, pp. 1744–1748, 2019, doi: 10.1109/ICTAI.2019.00259.
[29] F. Devriendt, J. Van Belle, T. Guns, and W. Verbeke, "Learning to rank for uplift modeling," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 10, pp. 4888–4904, 2022, doi: 10.1109/TKDE.2020.3048510.
[30] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.
[31] L. Breiman, "Random forests. Random Forests, 1–122," *Machine Learning*, vol. 45, no. 45, pp. 5–32, 2001.

## BIOGRAPHIES OF AUTHORS

**Risyda Miftahur Rahmah** is currently pursuing a Master's Degree in Artificial Intelligence at Gadjah Mada University (UGM), Yogyakarta, Indonesia. Her research interests are data science and artificial intelligence. She can be contacted at email: risyda.miftahur.r@mail.ugm.ac.id.

**Sigit Priyanta** received his B.Sc. (2000), M.Cs. (2004), and Ph.D. (2016) in Computer Science from Universitas Gadjah Mada (UGM), Indonesia. He is currently an Assistant Professor in the Department of Computer Science and Electronics at UGM, where he also serves as the Secretary of the Directorate of Education and Learning and is a member of the Software and Data Engineering Lab. His research interests include data mining, data science, big data, and mobile applications. He can be contacted at email: seagatejogja@ugm.ac.id.