Kyoung-Sook MOON, PhD
ksmoon@gachon.ac.kr
Gachon University, Gyeonggi, South Korea

Deokhyeon HAM, Master's student
hamch6910@korea.ac.kr
Korea University, Seoul, South Korea

Yein KWON, PhD student
yeinkwon@korea.ac.kr
Korea University, Seoul, South Korea

Hongjoong KIM, PhD (corresponding author)
hongjoong@korea.ac.kr
Korea University, Seoul, South Korea

An Adaptive Sampling Framework for Robust Anomaly Detection in Overlapping and Imbalanced Datasets

Abstract. *This study introduces an adaptive sampling framework for robust anomaly detection in overlapping and imbalanced datasets. The proposed approach integrates geometric oversampling with ensemble-based undersampling to overcome the limitations of conventional resampling techniques. Unlike interpolation-based methods such as SMOTE, the geometric oversampling strategy preserves the local structure of minority class data and minimises the generation of ambiguous synthetic samples. In parallel, the ensemble-based undersampling mechanism selectively retains majority instances to balance recall and precision, leading to consistent improvements in the F1-score across classifiers. Feature selection is incorporated to enhance model efficiency and interpretability by eliminating redundant variables and emphasising informative predictors. Experiments on the UCI Default of Credit Card Clients dataset, along with additional validation on the UCI Bank Marketing dataset, demonstrate that the framework achieves superior performance and generalisability under severe class imbalance and overlapping boundaries. From an economic cybernetics perspective, the proposed method strengthens decision reliability in high-stakes domains such as credit risk assessment, fraud detection, and cybersecurity.*

Keywords: *imbalanced data, anomaly detection, adaptive sampling, geometric oversampling, ensemble-based undersampling, precision–recall trade-off, economic cybernetics.*

JEL Classification: G17, C53, C55.

Received: 4 Nov 2025	**Revised: 2 June 2026**	**Accepted: 8 June 2026**

DOI: 10.24818/18423264/60.2.26.02

1. Introduction

This study addresses the challenge of detecting abnormal patterns in highly imbalanced and overlapping datasets, focusing on their relevance to economic systems and cybernetic decision-making. Anomaly detection plays a crucial role in identifying rare but high-impact events in financial systems. These anomalies, such as credit defaults, fraudulent transactions, or abnormal payment behaviours, often constitute only a small fraction of all observations, yet their accurate detection is vital for mitigating financial losses and maintaining market stability (Chandola et al., 2009; Ahmed et al., 2016; Coser et al., 2019).

A major obstacle in this task is the severe class imbalance inherent in financial data, where normal instances vastly outnumber abnormal ones. Under such conditions, conventional classifiers tend to bias toward the majority class, achieving high overall accuracy while failing to detect rare but crucial events. The challenge becomes even greater when class boundaries overlap, which is frequently observed in financial transactions where legitimate and anomalous behaviours share similar patterns. This overlap complicates the discrimination process, causing traditional classification methods to produce unstable decision boundaries and unreliable predictions (Lee and Kim, 2018).

To address these challenges, recent advances in learning from imbalanced data can be broadly categorised into data-level and algorithm-level approaches. Data-level approaches modify the dataset distribution through resampling, either by oversampling the minority class, undersampling the majority class, or combining both into hybrid strategies. Oversampling methods, such as the Synthetic Minority Oversampling Technique (SMOTE) (Chawla et al., 2002) and its extensions (Han et al., 2005; He and Garcia 2009; Liang et al., 2020), generate synthetic minority samples to alleviate imbalance. However, these interpolation-based techniques may distort the geometric structure of minority clusters and produce ambiguous instances, especially in overlapping regions. Conversely, undersampling methods such as random undersampling (Liu et al., 2009; Lin et al., 2017) reduce the majority class size, but risk losing informative samples, which can degrade generalisation performance. As a result, finding a proper trade-off between information retention and balanced learning remains a critical issue in imbalanced anomaly detection. In contrast, algorithm-level methods adjust the model's learning objective using cost-sensitive mechanisms, assigning higher penalties to the misclassification of minority instances (Nguyen et al., 2010; Kim et al., 2016; Muthukumaran et al., 2025). Although cost-sensitive learning provides a principled solution, it often relies on predefined cost parameters that are difficult to estimate in real-world financial applications. Therefore, adaptive and data-driven sampling mechanisms remain essential for robust and generalisable learning under severe imbalance.

Building on this perspective, the present study proposes an adaptive sampling framework that integrates geometric oversampling with ensemble-based undersampling to enhance anomaly detection in overlapping and imbalanced datasets. The geometric oversampling component generates synthetic samples along the intrinsic manifold of minority data, thereby preserving local structure and reducing the risk of unrealistic sample generation. In parallel, the ensemble-based undersampling mechanism selectively retains informative majority instances across multiple base learners, ensuring balanced learning without discarding essential information. This hybrid strategy enhances model robustness under extreme imbalance and overlapping distributions.

To further improve efficiency and interpretability, the framework incorporates feature selection to remove redundant or irrelevant attributes and emphasise informative predictors. High-dimensional financial datasets often contain noisy or correlated features that obscure minority patterns. By integrating feature selection with adaptive sampling, the proposed approach achieves both computational efficiency and improved discriminative power. Empirical experiments using the UCI Default of Credit Card Clients dataset (Yeh, 2009) and subsequent validation on the UCI Bank Marketing dataset (Moro et al., 2014) demonstrate that the proposed method consistently outperforms traditional resampling techniques, yielding higher evaluation metrics such as recall, precision, and F1-scores across multiple classifiers.

From an economic cybernetics perspective, the proposed methodology enhances decision reliability in systems characterised by asymmetric information and rare but high-impact risks. By reducing false negatives and preserving critical decision information, it supports more stable and informed decision-making in financial domains such as credit risk assessment, loan default prediction, and fraud detection. Overall, this study contributes a scalable and data-driven solution for achieving robust anomaly detection in overlapping and imbalanced datasets.

The remainder of this paper is organised as follows. Section 2 reviews the relevant literature on anomaly detection under imbalanced or overlapping data conditions. Section 3 describes the UCI Default of Credit Card Clients and UCI Bank Marketing datasets, summarising their key characteristics. Section 4 presents the proposed adaptive sampling framework, detailing its methodological components and experimental setup. Section 5 defines the evaluation metrics and reports the experimental results, highlighting the performance improvements achieved by the proposed approach. Finally, Section 6 concludes the paper by summarising the main findings and outlining possible directions for future research.

2. Related Work

Anomaly and fraud detection in financial systems has received extensive attention due to the asymmetric nature of risks and the rarity of abnormal events (Bolton and Hand, 2002; Phua et al., 2010). Conventional classification algorithms, when applied to highly imbalanced data, tend to favour the majority (normal) class, resulting in poor detection of critical minority cases. To overcome this limitation, the literature has developed two major categories of techniques: data-level resampling and algorithmic-level cost adaptation (He and Garcia, 2009; Fernández et al., 2018).

2.1 Data-Level Approaches

Data-level methods attempt to rebalance the dataset before training by adjusting the distribution of minority and majority samples. One of the most influential algorithms is the Synthetic Minority Oversampling Technique (SMOTE), which generates artificial minority instances through linear interpolation between existing samples (Chawla et al., 2002). Although effective in increasing minority representation, SMOTE tends to create samples in overlapping regions, which can blur class boundaries and reduce classifier specificity.

Subsequent variants such as Borderline-SMOTE (Han et al., 2005) and Safe-Level-SMOTE (Bunkhumpornpat et al., 2012) attempted to mitigate this issue by emphasising sample generation near decision boundaries or in low-risk areas of the feature space. ADASYN (He et al., 2008) further improved adaptability by generating more synthetic points in regions with higher learning difficulty. Despite these advancements, interpolation-based methods may still fail in datasets where minority and majority classes are highly overlapping, an issue common in financial transaction data, where legitimate and fraudulent behaviours exhibit similar statistical profiles.

Another line of research combines oversampling with undersampling to create hybrid strategies that balance diversity and representativeness. Examples include SMOTE-Tomek Links and SMOTE-ENN, which remove ambiguous samples after oversampling (He and Garcia, 2009; Pozzolo et al., 2015). While these hybrid approaches improve boundary clarity, they can be sensitive to noise and lack adaptivity across classifiers. Moreover, improper resampling during cross-validation can lead to data leakage, causing inflated performance metrics that fail to generalise to unseen data (Sáez et al., 2015).

2.2 Algorithmic-Level Approaches

Algorithmic-level solutions modify the learning objective rather than the data distribution. Cost-sensitive learning (Nguyen et al., 2010; Kim et al., 2016) assigns higher misclassification penalties to the minority class, encouraging the classifier to focus on rare events. Similarly, class weighting and threshold adjustment strategies have been widely used in credit scoring and fraud detection applications (Verbraken et al., 2014; Bahnsen et al., 2015). However, these approaches require accurate estimation of cost ratios, which are often unknown or unstable in real financial systems.

Another influential family of methods involves ensemble-based learning, which trains multiple base classifiers on different resampled subsets to enhance generalisation (Liu et al., 2009; Galar et al., 2012). These ensemble strategies outperform single resampling schemes by integrating diverse decision boundaries but remain sensitive to noise and boundary overlap in the training data. Consequently, many ensemble-based models still exhibit precision–recall instability, where improving recall comes at the expense of precision, limiting their usefulness in operational environments that demand both accuracy and reliability (Prati et al., 2015).

2.3 Limitations in Overlapping Financial Data

The UCI Default of Credit Card Clients dataset exemplifies a particularly challenging case of class overlap (Yeh and Lien, 2009). Normal and defaulting customers exhibit similar financial profiles, such as balance ratios and payment histories, making class separation non-trivial. As a result, tuning thresholds often improves recall at the cost of precision, with little net gain in the harmonic mean (F1-score). Many published studies using this dataset report F1-scores below 0.5, reflecting the inherent difficulty of the task (Gicic et al., 2024; Bhandary and Ghosh, 2025). Furthermore, some studies achieving higher F1 values suffer from methodological flaws, including oversampling applied before data partitioning, leading to optimistic but invalid results (Dal Pozzolo et al., 2014; Sáez et al., 2016).

2.4 Contribution of This Study

To overcome these limitations, the present study introduces a geometry-aware oversampling technique integrated within an ensemble-based undersampling framework. Unlike interpolation-based approaches that disregard the geometric structure of minority clusters, the proposed method preserves local topology and prevents artificial overlap between normal and abnormal samples. Combined with feature selection, this design maintains generalisation across classifiers while ensuring robustness in overlapping financial data. In doing so, this work extends the literature by providing a principled and reproducible framework for imbalance-aware learning in complex economic datasets.

3. Data Description

This study evaluates the proposed adaptive sampling framework using two distinct real-world financial datasets to ensure the robustness and generalisability of the model across different domains and time periods. The primary case study employs the widely used UCI Default of Credit Card Clients dataset (Yeh, 2009), which reflects real-world credit behaviour patterns observed in consumer lending. Furthermore, to validate the framework's effectiveness in a more recent and diverse context, the UCI Bank Marketing dataset (Moro et al., 2014) is introduced for additional empirical validation. Both datasets represent practically important cases in financial risk management and exhibit highly imbalanced target variable distributions. This inherent imbalance poses significant challenges for predictive modelling and directly motivates the sampling strategies explored in this study. The specific characteristics and preprocessing steps for each dataset are detailed in the following subsections.

3.1 Primary Case Study: UCI Default of Credit Card Clients Dataset

3.1.1 Dataset Overview

As the primary case study, this research utilises the Default of Credit Card Clients dataset, which consists of 25 financial and demographic variables, including repayment status, historical bill amounts, and past payment behaviour. Table 1 summarises these variables, where *default.payment.next.month* is defined as the binary target variable to indicate the default status. Variables with a small number of unique values such as *SEX, EDUCATION, MARRIAGE, PAY_0, PAY_2, PAY_3, PAY_4, PAY_5,* and *PAY_6* are treated as categorical features. The variable *ID*, which uniquely identifies each client but offers no predictive value, is removed. All remaining variables are treated as numerical features.

Table 1. Variables in the dataset

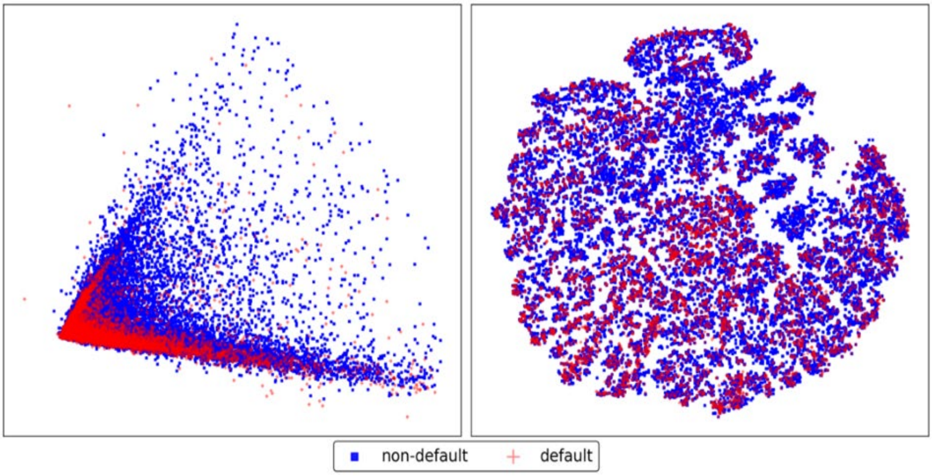
Variable	Meaning	Number of distinct values
ID	ID of each client	30000
LIMIT_BAL	Amount of given credit in NT dollars	81
SEX	Gender (1=male, 2=female)	2
EDUCATION	(1=graduate school, 2=university, 3=high school, 4=others, 5=unknown, 6=unknown)	7
MARRIAGE	Marital status (1=married, 2=single, 3=others)	4
AGE	Age in years	56
PAY_0	Repayment status in September, 2005 (-2=no consumption, -1=pay duly, 0=the use of revolving credit, 1=payment delay for one month, 2=payment delay for two months, ... 8=payment delay for eight months, 9=payment delay for nine months and above)	11
PAY_2	Repayment status in August, 2005 (scale same as above)	11
PAY_3	Repayment status in July, 2005 (scale same as above)	11
PAY_4	Repayment status in June, 2005 (scale same as above)	11
PAY_5	Repayment status in May, 2005 (scale same as above)	10
PAY_6	Repayment status in April, 2005 (scale same as above)	10
BILL_AMT1	Amount of bill statement in September, 2005 (NT dollar)	22723
BILL_AMT2	Amount of bill statement in August, 2005 (NT dollar)	22346
BILL_AMT3	Amount of bill statement in July, 2005 (NT dollar)	22026
BILL_AMT4	Amount of bill statement in June, 2005 (NT dollar)	21548
BILL_AMT5	Amount of bill statement in May, 2005 (NT dollar)	21010
BILL_AMT6	Amount of bill statement in April, 2005 (NT dollar)	20604
PAY_AMT1	Amount of previous payment in September, 2005 (NT dollar)	7943
PAY_AMT2	Amount of previous payment in August, 2005 (NT dollar)	7899
PAY_AMT3	Amount of previous payment in July, 2005 (NT dollar)	7518
PAY_AMT4	Amount of previous payment in June, 2005 (NT dollar)	6937
PAY_AMT5	Amount of previous payment in May, 2005 (NT dollar)	6897
PAY_AMT6	Amount of previous payment in April, 2005 (NT dollar)	6939
default.payment.next.month	Default payment (1=yes, 0=no)	2

Source: UCI Default of Credit Card Clients dataset (Yeh, 2009).

3.1.2 Class Imbalance and Data Overlap Analysis

This dataset presents a financially meaningful example of an imbalanced classification problem, where normal (non-default) observations account for approximately 80% of the data. Conventional classification methods that emphasise overall correctness tend to be biased toward the majority class, often overlooking rare but economically significant default cases. In financial risk analysis and fraud detection, overlooking such minority cases is much more costly than occasionally classifying a normal case as abnormal. Therefore, methods that increase the ability to detect minority cases without excessively sacrificing reliability are required, particularly under conditions where class distributions are skewed and decision errors impose asymmetric economic consequences.

In addition to class imbalance, this dataset presents another challenge: overlapping class distributions. A Principal Component Analysis (PCA) projection (Jolliffe, 2002) in Figure 1 (Left) reveals considerable overlap between normal (blue circle) and default (red plus) samples, indicating that the classes are not simply separable. A t-SNE visualisation (var der Maaten and Hinton, 2008) in Figure 1 (Right) further confirms that the minority class is embedded within the majority distribution, making classification more difficult.



(a) PCA plot of the data (b) t-SNE plot of the data

Figure 1. Data visualisation plot using (a) PCA and (b) t-SNE

Source: Calculation by authors.

3.1.3 Feature Engineering

To enrich the feature space with financially meaningful information, additional features are engineered, including the ratio of each payment amount to the corresponding bill amount, the total and average amounts of historical bill statements, and the total and average payments over the same period.

Pairwise Pearson correlation coefficients (Rodgers and Nicewander, 1988) are then computed to identify multicollinearity among features. Highly correlated variables with absolute correlation coefficients above 0.8 are removed to prevent redundancy and reduce model complexity. In addition, records containing financially unrealistic or inconsistent values are excluded to improve data quality.

Feature importance analysis is conducted prior to model training to identify the variables that most strongly contribute to the detection of default behaviour. Univariate feature selection is employed as a statistical filtering method that evaluates each feature individually based on its relevance to the target variable. This approach relies on univariate statistical criteria such as chi-square, ANOVA F-value, mutual information, or the Fisher score. The Fisher score, in particular, measures the discriminative power of a feature by comparing between-class variance to within-class variance. Features with higher Fisher scores exhibit stronger separation between classes and are therefore considered more informative for classification. Table 2 presents the Fisher scores for all features, ranked in descending order of discriminative power. As shown in Figure 2, the Fisher score distribution exhibits a sharp decline after the fourth feature, indicating that information relevant to class separation is highly concentrated in a small subset of variables. A moderate decline is observed from the fifth to the eighth feature, while the remaining features contribute marginally to class discrimination. Based on this evidence, three experimental configurations were defined: (i) top 4 features, (ii) top 8 features, and (iii) the full feature set. This design enables analysis of the trade-off between model efficiency and predictive performance under varying levels of feature relevance.

Table 2. Fisher scores of all features in the dataset and three configurations

Feature	Fisher Score	Configurations		
		All	Top 8	Top 4
PAY_0	2879.27	O	O	O
PAY_2	1783.67	O	O	O
PAY_3	1425.79	O	O	O
PAY_4	1187.51	O	O	O
LIMIT_BAL	570.00	O	O	-
SUM_PAYMENTS	249.57	O	O	-
PAY_AMT1	123.78	O	O	-
PAY_AMT2	104.17	O	O	-
PAY_AMT3	78.59	O	-	-
PAY_AMT6	73.90	O	-	-
PAY_AMT5	70.69	O	-	-
PAY_AMT4	67.56	O	-	-
SEX	50.34	O	-	-
PAY4_OVER_BILL5	46.02	O	-	-
EDUCATION	22.40	O	-	-
MARRIAGE	16.06	O	-	-
BILL_AMT1	14.68	O	-	-
AGE	1.92	O	-	-
PAY2_OVER_BILL3	0.86	O	-	-
PAY1_OVER_BILL2	0.61	O	-	-
PAY5_OVER_BILL6	0.16	O	-	-

Source: Calculation by authors.

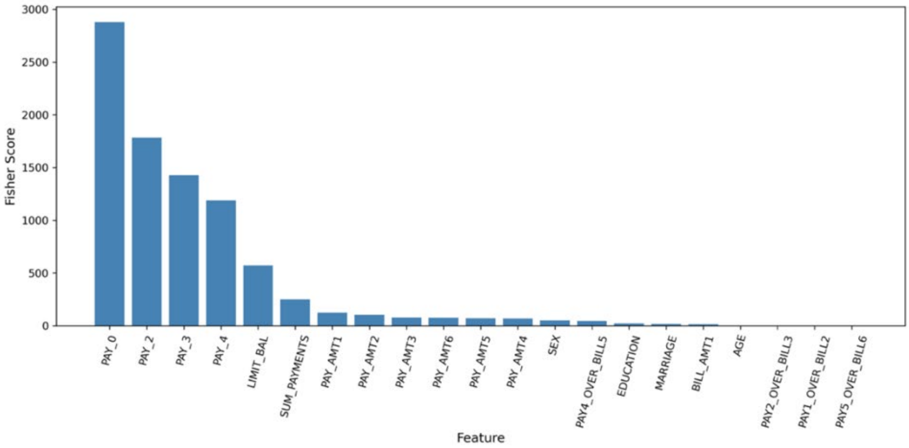


Figure 2. Fisher score ranking of input features
Source: Calculation by authors.

3.2 Additional Validation: UCI Bank Marketing Dataset

To validate the proposed framework on a more recent dataset from a different financial domain, this study additionally analyses the UCI Bank Marketing dataset. This dataset contains records from direct marketing campaigns of a Portuguese bank, where the classification goal is to predict whether a client will subscribe to a term deposit.

The dataset consists of 45,211 instances with a mix of numerical and categorical variables related to client demographics and campaign details. Similar to the primary case dataset, it exhibits severe class imbalance. The minority class (subscribers, labeled as 'yes') accounts for only 11.7% (5,289 instances) of the data, while the majority class ('no') makes up 88.3% (39,922 instances). This high imbalance and overlapping feature spaces make it highly suitable for testing the resampling framework.

For a consistent comparative analysis, the same preprocessing steps used in the primary case study were applied. Furthermore, following the procedure in Section 3.1.3, feature selection was performed based on Fisher scores. Rather than using all available variables, the top-ranked features were selected to maintain computational efficiency. The evaluation results for this dataset are discussed in Section 5.3.

4. Methods

4.1 Oversampling and Undersampling

The Synthetic Minority Oversampling Technique (SMOTE) is a widely used oversampling approach designed to address class imbalance in classification problems. Instead of simply replicating minority class instances, SMOTE generates synthetic samples by interpolating between each minority instance and its nearest minority neighbor in feature space. This helps mitigate overfitting that may occur with random oversampling and provides a smoother decision region for the minority class. Unlike simple reweighting methods that increase the importance of existing minority samples, SMOTE creates additional data points and modifies the distribution of the training set. However, these synthetic samples are constrained to lie within the convex hull of the original minority instances. As a result, while SMOTE enhances minority class representation, it may not sufficiently expand the data manifold in complex or overlapping class distributions.

In this study, we propose a novel oversampling technique called *GEometric OverSampling* (GEOS), designed to generate synthetic samples for the minority class based on geometric constraints in feature space. Let x be an arbitrary minority (abnormal) instance as in Figure 3. Denote by n_x the nearest majority (normal) instance to x, and let their distance be $d_x = \|x - n_x\|$. We define a ball B_x in the feature space centered at x with radius r_x, where $0 < r_x < d_x$. GEOS generates a synthetic sample s_x by randomly sampling a point within B_x. Numerical feature values of s_x are generated within this geometric region, while the categorical features are directly copied from x to preserve semantic validity.

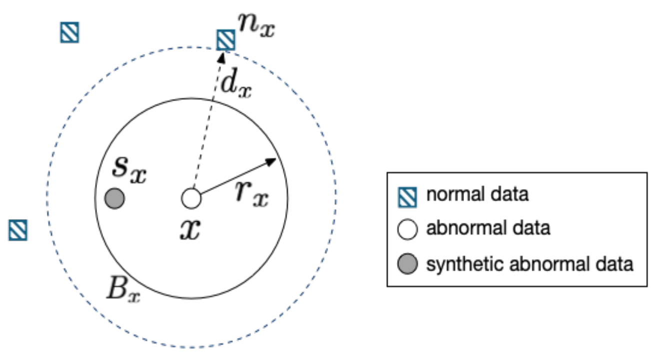
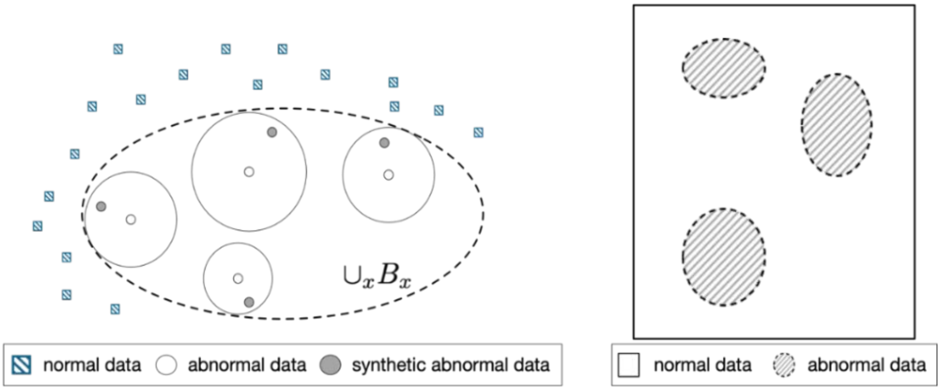


Figure 3. Illustration of GEOS method
Source: Illustration by authors.

The procedure is repeated for multiple minority instances to generate a richer and more representative minority region, as illustrated in Figure 4 (Left). The union of the local geometric regions $\cup_x B_x$ forms an estimated minority data manifold, resulting in a clearer separation between the minority region (original and synthetic abnormal data) and the majority region (normal data), as shown in Figure 4 (Right). Unlike SMOTE, which creates synthetic samples by linear interpolation between existing minority points and, therefore, does not expand the minority data manifold, the proposed GEOS method produces genuinely new and previously unseen samples. As a result, it provides a better approximation of the minority class boundary and enhances minority class representation in complex feature spaces.



(a) Local balls around abnormal samples and their union. (b) Resulting minority regions after oversampling.

Figure 4. GEOS and class-region formation

Source: Illustration by authors.

Preliminary experiments indicate that when the dataset is highly imbalanced or exhibits substantial class overlap, the effectiveness of oversampling alone becomes limited. To improve classification performance under such conditions, undersampling techniques can be incorporated in combination with oversampling. One commonly used method is Edited Nearest Neighbours (ENN), a data-cleaning technique that removes majority or minority instances whose class labels differ from the majority of their k-nearest neighbours. By eliminating noisy or ambiguous samples located near class boundaries, ENN reduces class overlap and refines the decision boundary, thereby improving model generalisation. ENN is often used alongside SMOTE to simultaneously enhance minority class representation and suppress noise in the feature space.

Another effective strategy is an ensemble-based undersampling. Rather than attempting to generate additional synthetic minority samples, which may have a limited impact when abnormal data is extremely sparse, this approach reduces the size of the majority class. Specifically, instead of using all majority instances, a subset is selected to balance the class distribution, and a classifier is trained on this reduced dataset. The process is repeated multiple times to build an ensemble of classifiers, and final predictions are obtained via majority voting. Since the proposed GEOS method generates synthetic samples within a localised region around each minority instance, the marginal benefit of ENN becomes limited due to reduced boundary noise. Therefore, this study adopts the ensemble-based undersampling approach. The undersampling ratio (α) is treated as a tunable hyperparameter and is defined as the proportion of majority class instances retained relative to their original size. An undersampling ratio of 1.0 indicates that no undersampling is applied. Meanwhile, oversampling is applied until the minority and majority classes are equal in size.

4.2 Adaptive Sampling Strategy

To address severe class imbalance while preserving the informative majority class structure, this study adopts an adaptive sampling framework as in Figure 5 that integrates GEOS with ensemble-based undersampling.

First, GEOS is applied to generate synthetic minority class instances within localised feature regions, thereby reinforcing minority representation without introducing boundary noise. Next, multiple balanced training subsets are constructed by randomly drawing subsets from the majority class, each equal in size to the oversampled minority class.

A separate classifier is trained on each balanced subset and its corresponding minority class, forming an undersampling ensemble. During inference, predictions from all ensemble members are aggregated using majority voting to obtain the final classification decision. This adaptive strategy reduces sampling bias, preserves decision boundary integrity, and mitigates overfitting to duplicated minority samples.

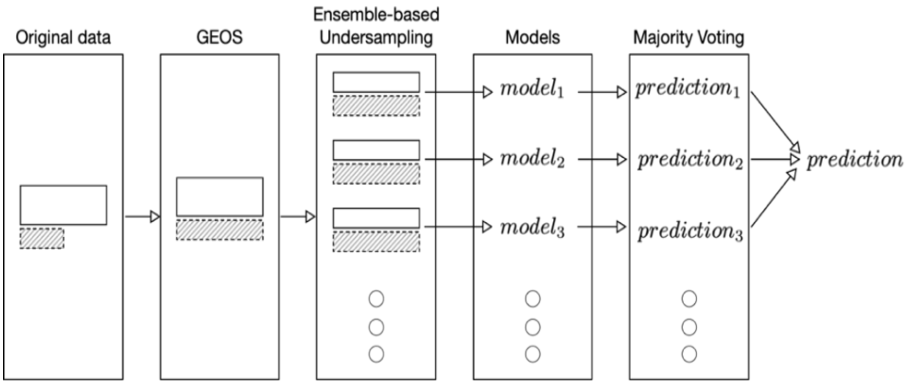


Figure 5. The algorithm of an adaptive sampling approach
Source: Illustration by authors.

The highlights of the proposed method are as follows:

First, the proposed GEOS creates new synthetic samples within nonlinear, safe regions around minority instances, preventing intrusion into majority-class areas.

Secondly, an ensemble-based undersampling scheme randomly selects subsets of the majority class to form multiple balanced training datasets, with which the majority voting eventually improves the prediction performance.

Thirdly, this two-step method defines a surface within the feature space separating majority and minority classes, ensuring better coverage of minority instances while preserving essential majority patterns.

Lastly, as shown in the next section, the proposed method is not limited to a specific machine learning model, but it can be applied to prediction models in general to improve their results. In particular, it identifies the minority class better than other approaches.

5. Empirical Results and Discussion

5.1 Evaluation Metrics

Model performance is evaluated using four standard classification measures: accuracy, precision, recall, and F1-score. Accuracy indicates the proportion of correctly classified samples among all observations, offering a general measure of correctness:

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

where TP (true positives) and TN (true negatives) denote correctly identified positive and negative cases, respectively, while FP (false positives) are negative cases incorrectly labelled as positive and FN (false negatives) are positive cases incorrectly classified as negative.

Precision measures the reliability of positive predictions, that is, the proportion of correctly predicted positive cases among all predicted positives, while recall (also called sensitivity) quantifies the model's ability to detect positive cases, that is, the proportion of actual positive cases that are successfully identified:

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}$$

The F1-score is the harmonic mean of precision and recall and provides a balanced evaluation when both false positives and false negatives are of concern:

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

In imbalanced classification problems such as in our dataset where default cases are significantly outnumbered by non-default cases (Section 4.2), accuracy alone is often misleading because a classifier may achieve high accuracy simply by predicting the majority class. In financial risk assessment, misclassifying default cases as non-default (false negatives) entails high economic cost, making recall particularly important. However, recall cannot be increased without affecting precision due to their intrinsic trade-off: a lower decision threshold increases the detection of positive cases but introduces additional false alarms, whereas a higher threshold reduces false positives at the risk of missing true positives. For this reason, the F1-score is employed as another principal metric in this study, as it balances minority class detection with classification reliability.

5.2 Experimental Findings on the UCI Default of Credit Card Clients Dataset

The experimental analysis evaluates the impact of three methodological dimensions, sampling strategy, feature selection, and classifier choice, on predictive performance. The results are compared across five dataset configurations (Raw, SMOTE, GEOS, GEOS8, and GEOS4) and multiple undersampling ratios using accuracy, precision, recall, and F1-score. Raw refers to the dataset without oversampling. SMOTE augments the minority class by generating synthetic samples until the class sizes are equal. GEOS applies the proposed geometric oversampling method as an alternative to SMOTE. GEOS8 and GEOS4 denote GEOS variants using the top 8 and top 4 features, respectively, selected based on Fisher scores. Eight undersampling ratios from 0.3 to 1.0 are considered. Five benchmark classifiers are used for comparison: Logistic Regression (LR), AdaBoost (AB), Random Forest (RF), XGBoost (XGB), and LightGBM (LGBM). A summary of the results is provided in Table 3, and graphical trends are presented throughout this section.

Table 3. Summary of accuracy (A), precision (P), recall (R), and F1-score (F1) across classifiers, dataset configurations, and undersampling ratios

Classifier	Dataset	Undersampling ratio (α)											
		1.0				0.5				0.3			
		A	P	R	F1	A	P	R	F1	A	P	R	F1
LR	Raw	0.808	0.683	0.270	0.388	0.807	0.593	0.457	0.516	0.753	0.462	0.604	0.524
	SMOTE	0.715	0.412	0.631	0.499	0.732	0.432	0.619	0.509	0.742	0.446	0.614	0.516
	GEOS	0.723	0.421	0.621	0.502	0.744	0.449	0.613	0.518	0.745	0.450	0.614	0.519
	GEOS8	0.721	0.418	0.623	0.501	0.755	0.465	0.600	0.524	0.745	0.450	0.614	0.519
	GEOS4	0.727	0.426	0.621	0.505	0.755	0.464	0.599	0.523	0.745	0.450	0.615	0.520
AB	Raw	0.821	0.677	0.392	0.496	0.811	0.597	0.482	0.534	0.775	0.498	0.600	0.544
	SMOTE	0.768	0.486	0.560	0.520	0.770	0.491	0.639	0.556	0.768	0.487	0.607	0.540
	GEOS	0.770	0.490	0.558	0.522	0.755	0.467	0.626	0.535	0.770	0.490	0.609	0.543
	GEOS8	0.746	0.452	0.620	0.523	0.750	0.462	0.670	0.547	0.769	0.488	0.610	0.542
	GEOS4	0.755	0.469	0.662	0.549	0.765	0.482	0.615	0.541	0.779	0.507	0.588	0.544
RF	Raw	0.819	0.664	0.393	0.494	0.802	0.562	0.543	0.552	0.757	0.470	0.640	0.542
	SMOTE	0.792	0.537	0.537	0.537	0.773	0.496	0.610	0.547	0.756	0.468	0.650	0.545
	GEOS	0.798	0.552	0.525	0.538	0.778	0.504	0.600	0.548	0.751	0.462	0.645	0.538
	GEOS8	0.802	0.564	0.517	0.539	0.779	0.506	0.596	0.548	0.756	0.468	0.650	0.544
	GEOS4	0.801	0.561	0.524	0.542	0.782	0.513	0.590	0.549	0.754	0.466	0.647	0.542
XGB	Raw	0.813	0.641	0.384	0.480	0.799	0.552	0.546	0.549	0.738	0.444	0.658	0.530
	SMOTE	0.792	0.545	0.443	0.488	0.769	0.488	0.587	0.533	0.738	0.445	0.673	0.536
	GEOS	0.794	0.553	0.445	0.493	0.768	0.486	0.586	0.532	0.741	0.448	0.667	0.536
	GEOS8	0.788	0.530	0.480	0.504	0.767	0.484	0.574	0.525	0.734	0.440	0.666	0.530
	GEOS4	0.788	0.530	0.475	0.501	0.774	0.497	0.585	0.537	0.736	0.442	0.666	0.532
LGBM	Raw	0.822	0.682	0.385	0.492	0.807	0.576	0.531	0.552	0.762	0.477	0.635	0.545
	SMOTE	0.810	0.594	0.488	0.536	0.784	0.517	0.597	0.554	0.758	0.472	0.651	0.548
	GEOS	0.807	0.587	0.477	0.526	0.785	0.517	0.601	0.556	0.759	0.473	0.653	0.548
	GEOS8	0.769	0.489	0.590	0.535	0.767	0.486	0.636	0.551	0.760	0.475	0.650	0.549
	GEOS4	0.767	0.485	0.632	0.549	0.765	0.482	0.641	0.550	0.759	0.474	0.653	0.549

Source: Calculation by authors.

5.2.1 Effect of Oversampling on Minority Class Detection

Baseline models trained on the Raw dataset show relatively high accuracy and precision but suffer from low recall, confirming the presence of class imbalance and minority under-representation. When oversampling is applied, both SMOTE and GEOS notably improve recall across all classifiers and undersampling ratios. This demonstrates that oversampling substantially enhances the detection of minority-class observations corresponding to abnormal financial behaviour.

Figure 6 shows the recall (Left) and F1-score (Right) when the undersampling ratio is 1.0 (no undersampling) for each oversampling configuration, reflecting the overall predictive strength of each method. Both SMOTE and GEOS consistently outperform the Raw dataset in both recall and F1-score, verifying the benefit of minority-class augmentation.

Table 4 presents detailed F1-scores across classifiers and datasets and the relative F1 improvement compared to Raw, computed by

$$F1_{improvement} = \left(\frac{F1_{dataset} - F1_{Raw}}{F1_{Raw}}\right) \times 100(\%)$$

for each dataset configuration, corroborating the quantitative magnitude of these improvements. On average, recall increases from 0.3649 (Raw) to 0.5316 (SMOTE) and 0.5829 (GEOS4), indicating that geometric oversampling yields a greater improvement in minority detection than traditional SMOTE. In terms of overall F1 performance, GEOS4 achieves the highest average (0.5292), followed by GEOS8 (0.5202), GEOS (0.5161), SMOTE (0.5160), and Raw (0.4701).

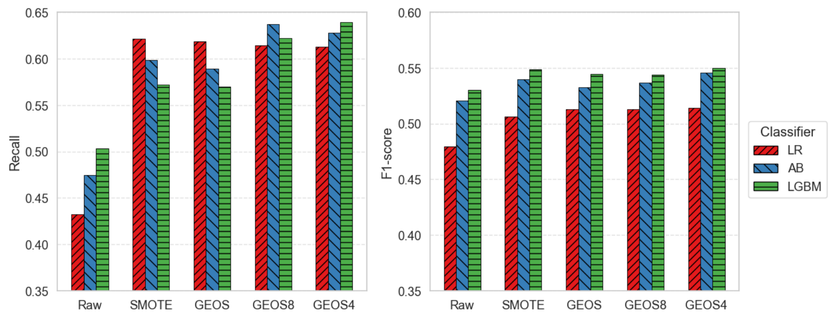


Figure 6. Comparison of (Left) recall and (Right) F1 across oversampling dataset configurations (no undersampling)

Source: Calculation by authors.

Table 4. Quantitative summary of oversampling performance (no undersampling)

Classifier	Dataset	F1	$F1_{improvement}$
LR	Raw	0.3875	
	SMOTE	0.4986	28.67%
	GEOS	0.5019	29.52%
	GEOS8	0.5005	29.16%
	GEOS4	0.5053	30.40%
AB	Raw	0.4962	
	SMOTE	0.5203	4.86%
	GEOS	0.5217	5.14%
	GEOS8	0.5229	5.38%
	GEOS4	0.5488	10.60%
RF	Raw	0.4940	
	SMOTE	0.5369	8.68%
	GEOS	0.5380	8.91%
	GEOS8	0.5393	9.17%

Classifier	Dataset	F1	$F1_{improvement}$
	GEOS4	0.5418	9.68%
XGB	Raw	0.4802	
	SMOTE	0.4884	1.71%
	GEOS	0.4930	2.67%
	GEOS8	0.5037	4.89%
	GEOS4	0.5010	4.33%
LGBM	Raw	0.4924	
	SMOTE	0.5357	8.79%
	GEOS	0.5260	6.82%
	GEOS8	0.5347	8.59%
	GEOS4	0.5489	11.47%
Mean	Raw	0.4701	
	SMOTE	0.5160	9.77%
	GEOS	0.5161	9.80%
	GEOS8	0.5202	10.67%
	GEOS4	0.5292	12.57%

Source: Calculation by authors.

Although SMOTE effectively increases recall, it often generates synthetic samples close to class boundaries, introducing ambiguity and reducing precision. In contrast, the proposed GEOS constructs synthetic data within localised hyperspherical regions that better preserve intrinsic minority-class geometry. This design mitigates boundary noise and improves class separability. Empirical results confirm that GEOS consistently outperforms SMOTE in F1-score, providing a more favourable balance between precision and recall. Among all variants, GEOS4 achieves the best performance and exhibits high robustness to undersampling variation.

The Precision–Recall curves in Figure 7 further illustrate how GEOS and its feature-selected variants (GEOS8, GEOS4) improve minority detection. Compared with SMOTE, GEOS4 (red solid line) consistently achieves higher precision at equivalent recall levels, confirming enhanced discrimination for rare events.

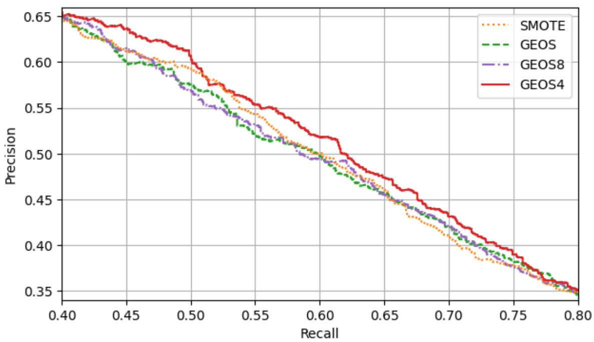


Figure 7. Precision–Recall curves for different dataset configurations using LGBM under an undersampling ratio of 1.0 (no undersampling)

Source: Calculation by authors.

5.2.2 Feature Selection Enhances Both Efficiency and Predictive Quality

To examine the impact of redundant variables, feature selection was incorporated via GEOS8 and GEOS4. Reducing feature dimensionality from the full set to the top 8 and top 4 features mitigates multicollinearity and noise while improving computational efficiency.

The results reveal that feature selection improves both recall and F1-score, most notably for GEOS4, indicating that compact, informative subsets strengthen minority representation. This aligns with the overall performance ranking (GEOS4 > GEOS8 > GEOS > SMOTE ≈ Raw), confirming that combining geometric oversampling with targeted feature selection yields the most stable outcomes.

As shown in Figure 8, recall and F1 progressively improve from GEOS to GEOS8 to GEOS4, suggesting that removing redundant features enhances classification capability. GEOS4 also reduces model training time by 22.56% (AB) and 13.73% (RF) while maintaining or improving predictive accuracy. These findings confirm that feature selection contributes to both robustness and efficiency, supporting its practicality in large-scale or real-time financial risk analytics.

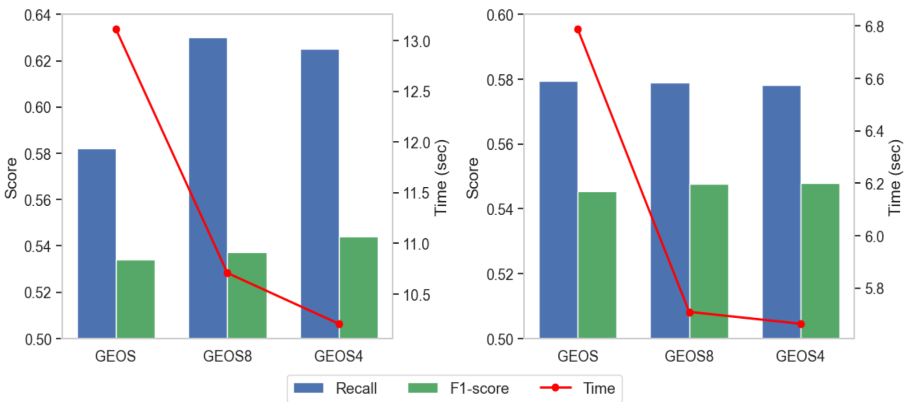


Figure 8. Comparison of the average recall, F1-score and computational time (sec) for GEOS, GEOS8 and GEOS4 for (Left) AB and (Right) RF

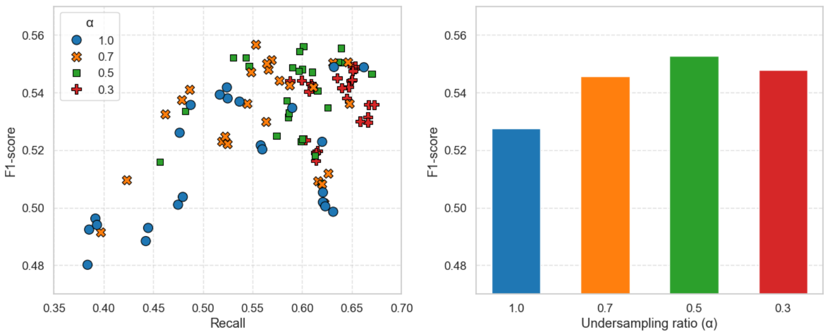
Source: Calculation by authors.

5.2.3 Role of Undersampling in Balancing Class Distributions

Undersampling was jointly applied with oversampling to balance class proportions and limit majority dominance. By varying the undersampling ratio (α = 0.3, 0.4, …, 1.0), we assessed the trade-off between preserving majority information and improving minority detectability.

Moderate undersampling (α = 0.5) produces the best overall outcomes, yielding strong recall and F1 without excessive false positives, whereas aggressive undersampling (α = 0.3) achieves the highest recall but reduces precision due to information loss. Overall averages show that smaller ratios tend to increase recall and F1, suggesting that heavier undersampling enhances minority recognition at the cost of precision.

GEOS4 exhibits a relatively stable F1 across all ratios, highlighting its robustness to undersampling variation. Figure 9 (Left) shows the variation of recall and F1-score across undersampling ratios, and Figure 9 (Right) shows F1-scores for several undersampling ratios. Table 5 presents representative confusion matrices showing how extreme undersampling inflates false alarms. Together, these results indicate that α = 0.5 provides the most favourable equilibrium between recall and precision, thereby enhancing classifier reliability.



(a) F1-score vs. Recall (b) Average F1-scores

Figure 9. Effect of different undersampling ratio $\alpha = 0.3, 0.5, 0.7, 1.0$.

Source: Calculation by authors.

Table 5. Confusion-matrix illustration for undersampling ratio α (LGBM + GEOS4)

Undersampling ratio (α)	TP FN	FP TN	Precision	Recall	F1-score
1.0	797 464	846 3507	0.4851	0.6320	0.5489
0.9	797 464	836 3517	0.4881	0.6320	0.5508
0.8	794 467	817 3536	0.4929	0.6297	0.5529
0.7	796 465	836 3517	0.4877	0.6312	0.5503
0.6	807 454	856 3497	0.4853	0.6400	0.5520
0.5	808 453	867 3486	0.4824	0.6408	0.5504
0.4	823 438	918 3435	0.4727	0.6527	0.5483
0.3	824 437	916 3437	0.4736	0.6534	0.5492

Source: Calculation by authors.

5.2.4 Classifier Comparison and Model Stability

The final analysis evaluates the classifier performance and stability under different sampling strategies. Figure 10 summarises recall and F1-scores for all five classifiers (LR, RF, AB, XGB, LGBM) across dataset variants, confirming that LGBM and RF deliver the most stable and accurate results, followed by XGB, AB, and LR. These findings highlight the importance of model choice: LGBM and RF achieve higher mean F1 and lower variability, indicating stronger generalisation under sampling perturbations.

LGBM, in particular, demonstrates the highest and most stable F1, reflecting its ability to capture nonlinear patterns with low sensitivity to sampling variation. In contrast, XGB shows higher variability in recall, while LR is less robust under imbalance.

The best-performing configuration involves GEOS4 oversampling dataset with an undersampling ratio of 0.5 and LightGBM classifier, offering the best trade-off between accuracy, efficiency, and stability. Figure 11 compares the classifier performance across undersampling ratios, and Table 6 consolidates the optimal configuration.

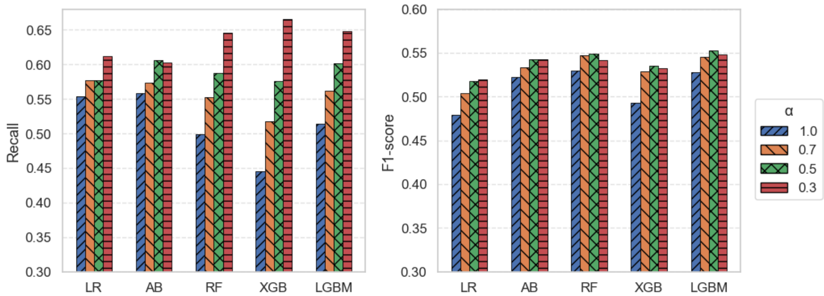


Figure 10. Model performance comparison across classifiers

Source: Calculation by authors.

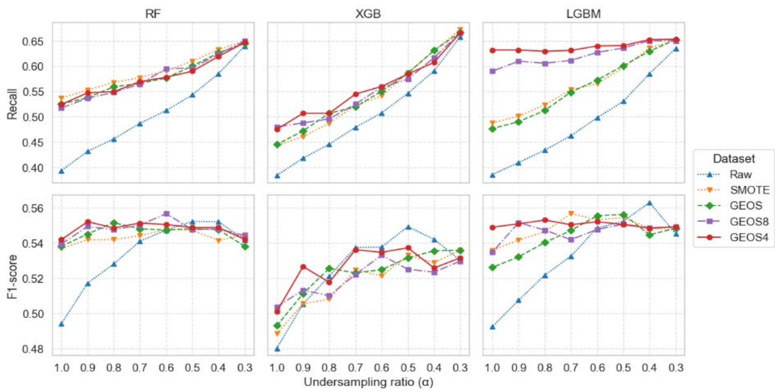


Figure 11. Classifier performance variation across undersampling ratios (α)

Source: Calculation by authors.

Table 6. Optimal hybrid sampling–classification configuration achieving the best trade-off between predictive accuracy and computational efficiency

Component	Selected Setting	Rationale
Oversampling dataset configuration	GEOS	Highest recall and F1 stability
Undersampling ratio	Ratio = 0.5	Balanced precision–recall
Feature selection	Top 4 (Fisher)	Faster training
Classifier	LGBM	Highest F1 with lowest variation

Source: Summary by authors.

Two main conclusions emerge: (1) the proposed adaptive sampling strategy significantly boosts recall and F1 without unacceptable precision loss; (2) these improvements are consistent across models, demonstrating generalisability.

This configuration effectively minimises false negatives, a key requirement in financial risk analytics such as credit default prediction, loan screening, and fraud detection, while maintaining computational feasibility for deployment.

5.3 Validation Results on the UCI Bank Marketing Dataset

To confirm the generalisability of the proposed framework, the experimental procedures were replicated using the UCI Bank Marketing dataset. Table 7 summarises the accuracy (A), precision (P), recall (R), and F1-score (F1) across the tested classifiers, varying dataset configurations, and undersampling ratios. The overall trends in the table align closely with the findings from the primary Default of Credit Card Clients dataset. While standard models struggle with the severe 11.7% minority class ratio, the proposed GEOS4 framework maintains competitive recall and F1-scores and significantly prevents the degradation of minority class detection. This consistency indicates that the framework's effectiveness is not limited to a specific dataset, but is applicable to different financial contexts.

Table 7. Summary of accuracy (A), precision (P), recall (R), and F1-score (F1) across classifiers, dataset configurations, and undersampling ratios

Classifier	Dataset	Undersampling ratio (α)											
		1.0				0.5				0.3			
		A	P	R	F1	A	P	R	F1	A	P	R	F1
LR	Raw	0.886	0.582	0.207	0.306	0.884	0.524	0.406	0.458	0.870	0.468	0.569	0.513
	SMOTE	0.803	0.359	0.804	0.496	0.803	0.359	0.805	0.496	0.803	0.359	0.808	0.497
	GEOS	0.802	0.358	0.811	0.497	0.802	0.358	0.808	0.496	0.801	0.356	0.807	0.494
	GEOS8	0.803	0.359	0.808	0.497	0.800	0.356	0.810	0.495	0.801	0.357	0.807	0.495
	GEOS4	0.802	0.358	0.809	0.496	0.802	0.358	0.805	0.495	0.800	0.355	0.808	0.493
AB	Raw	0.899	0.618	0.436	0.512	0.895	0.554	0.664	0.604	0.873	0.484	0.767	0.594
	SMOTE	0.887	0.526	0.642	0.578	0.882	0.508	0.768	0.611	0.864	0.464	0.815	0.591
	GEOS	0.884	0.517	0.641	0.572	0.876	0.490	0.780	0.602	0.860	0.456	0.818	0.585
	GEOS8	0.876	0.490	0.702	0.577	0.874	0.485	0.785	0.600	0.863	0.460	0.809	0.587
	GEOS4	0.834	0.410	0.864	0.556	0.830	0.404	0.868	0.552	0.830	0.405	0.875	0.554
RF	Raw	0.899	0.647	0.366	0.467	0.898	0.569	0.643	0.604	0.878	0.498	0.761	0.602
	SMOTE	0.889	0.529	0.705	0.604	0.873	0.483	0.799	0.602	0.851	0.439	0.836	0.576
	GEOS	0.891	0.539	0.647	0.588	0.872	0.481	0.786	0.597	0.850	0.436	0.836	0.573
	GEOS8	0.891	0.543	0.625	0.581	0.870	0.476	0.780	0.592	0.852	0.440	0.832	0.576

Classifier	Dataset	Undersampling ratio (α)											
		1.0				0.5				0.3			
		A	P	R	F1	A	P	R	F1	A	P	R	F1
	GEOS4	0.889	0.529	0.715	0.608	0.868	0.473	0.814	0.598	0.849	0.435	0.837	0.572
XGB	Raw	0.905	0.633	0.501	0.560	0.902	0.577	0.703	0.634	0.887	0.522	0.798	0.631
	SMOTE	0.901	0.596	0.554	0.574	0.894	0.544	0.730	0.624	0.879	0.499	0.822	0.621
	GEOS	0.901	0.597	0.546	0.571	0.895	0.549	0.731	0.627	0.878	0.497	0.810	0.616
	GEOS8	0.898	0.581	0.546	0.563	0.891	0.536	0.723	0.615	0.880	0.502	0.809	0.620
	GEOS4	0.893	0.545	0.678	0.604	0.886	0.519	0.783	0.624	0.871	0.481	0.821	0.606
LGBM	Raw	0.904	0.640	0.464	0.538	0.902	0.579	0.704	0.635	0.885	0.515	0.794	0.624
	SMOTE	0.895	0.550	0.690	0.612	0.890	0.529	0.778	0.630	0.875	0.489	0.829	0.615
	GEOS	0.900	0.573	0.664	0.615	0.886	0.519	0.784	0.624	0.869	0.476	0.841	0.608
	GEOS8	0.893	0.542	0.724	0.620	0.884	0.512	0.812	0.628	0.871	0.479	0.842	0.611
	GEOS4	0.853	0.445	0.885	0.593	0.852	0.443	0.876	0.588	0.848	0.436	0.883	0.584

Source: Calculation by authors.

Figure 12 further illustrates the relationship between the undersampling ratio and recall for the RF, XGB, and LGBM classifiers. Across all tested algorithms and varying ratios, the proposed GEOS4 method consistently achieves the highest recall values compared to the baseline configurations. This confirms that integrating geometric oversampling with ensemble-based undersampling effectively preserves the local structure of the minority class, allowing the models to correctly identify target instances (term deposit subscribers) even in highly imbalanced and overlapping data spaces.

Overall, the supplementary experiments on the bank marketing data validate the robustness of the GEOS4 framework, proving its reliability for real-world anomaly detection tasks.

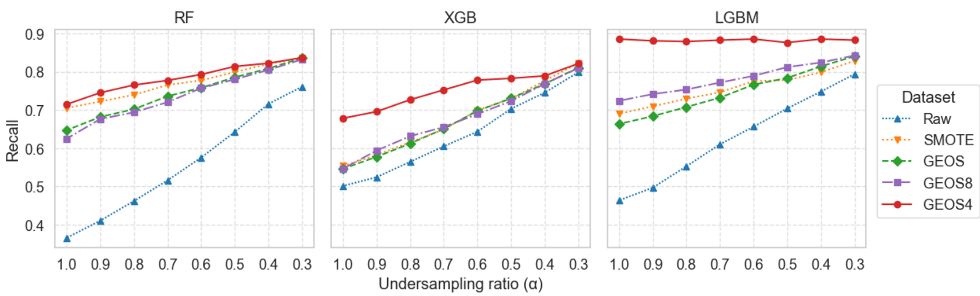


Figure 12. Classifier performance variation across undersampling ratios (α)

Source: Calculation by authors.

5.4 Discussion

The empirical findings offer meaningful implications from both economic and decision-theoretic perspectives. In domains such as credit risk assessment and insurance claim validation, false negatives misclassifying high-risk entities as low-risk incur significantly higher financial losses than false positives. By maximising minority class detection, the proposed framework directly mitigates these latent risk exposures.

Furthermore, real-world financial datasets, as demonstrated by both the credit default and bank marketing cases, frequently exhibit overlapping boundaries alongside severe class imbalance. Conventional techniques like SMOTE often generate synthetic samples along linear paths that cross class regions, which increases boundary ambiguity. In contrast, the proposed geometric oversampling restricts synthetic generation to localised regions. This preserves the original boundary structure and improves class separability, making the approach highly effective for detecting anomalies in complex nonlinear spaces, such as payment fraud and cyberattack monitoring.

Finally, integrating feature selection enhances economic efficiency by reducing computational overhead. By isolating only the most discriminative predictors, the framework improves accuracy while minimising training and deployment costs. This scalability confirms its viability for real-time decision automation in high-stakes financial and cyber-physical systems.

6. Conclusions

This study proposed an adaptive sampling framework that integrates geometric oversampling, ensemble-based undersampling, and feature selection to address severe class imbalance in anomaly detection. The method effectively enhances minority-class detection while maintaining computational efficiency, which is crucial for high-dimensional financial analytics.

The empirical results, **validated across both the primary default of credit card clients and the recent UCI bank marketing datasets**, highlight three main contributions. First, the geometric oversampling approach outperforms interpolation-based methods like SMOTE by generating synthetic samples within localised regions, preserving the minority-class structure, and reducing boundary ambiguity. Second, adopting a moderate undersampling ratio prevents majority-class dominance without losing critical decision information, leading to a strong precision-recall balance. Third, Fisher score-based feature selection eliminates redundant variables, improving both training efficiency and system scalability.

From a practical standpoint, the framework significantly reduces false negatives – a critical requirement in financial and security applications where undetected anomalies lead to severe losses. Its model-agnostic design ensures stable performance gains across multiple classifiers, with LightGBM achieving the best overall results under the proposed hybrid sampling configuration. **Furthermore, its consistent performance across different time periods and financial domains confirms its strong generalisability.**

Overall, this research provides a robust, interpretable, and computationally feasible sampling strategy for imbalanced learning in high-stakes environments. Future work will extend this framework by developing an online adaptive version for streaming data and concept drift, and by integrating explainable AI mechanisms to enhance transparency in economic decision systems.

References

[1] Ahmed, M., Mahmood, A.N, Islam, M.R. (2016), *A survey of anomaly detection techniques in financial domain. Future Generation Computer Systems*, 55, 278-288.

[2] Bahnsen, A.C., Aouada, D., Ottersten, B. (2014), *Example-dependent cost-sensitive logistic regression for credit scoring, 13th International Conference on Machine Learning and Applications*, MI, USA, 03-06 December 2014.

[3] Bhandary, R., Ghosh, B.K. (2025), *Credit Card Default Prediction: An Empirical Analysis on Predictive Performance Using Statistical and Machine Learning Methods. Journal of Risk and Financial Management*, 18(1), 23.

[4] Bolton, R.J., Hand, D.J. (2002), *Statistical fraud detection: a review. Statistical Science*, 17(3), 235-255.

[5] Bunkhumpornpat, C., Sinapiromsaran, K., Lursinsap, C. (2009), *Safe-Level-SMOTE: Safe-Level-Synthetic Minority Over-Sampling TEchnique for Handling the Class Imbalanced Problem. In: Theeramunkong, T., Kijsirikul, B., Cercone, N., Ho, TB. (eds) Advances in Knowledge Discovery and Data Mining. PAKDD 2009. Lecture Notes in Computer Science*, 5476. Springer, Berlin, Heidelberg, Germany.

[6] Chandola, V., Banerjee, A., Kumar, V. (2009), *Anomaly detection: A survey. ACM Computing Surveys*, 41(3), 1-58.

[7] Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P. (2002), *SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research*, 16, 321-357.

[8] Coşer, A., Maer-Matei, M.M., Albu, C. (2019). *Predictive models for loan default risk assessment. Economic Computation and Economic Cybernetics Studies and Research*, 53(2), 149-165.

[9] Dal Pozzolo, A., Caelen, O., Johnson, R.A., Bontempi, G. (2015), *Calibrating probability with undersampling for unbalanced classification, 2015 IEEE Symposium Series on Computational Intelligence*, CapeTown, South Africa, 07-10 December 2015.

[10] Fernandez, A., Garcia, S., Galar, M., Prati, R.C., Krawczyk, B. (2018*), Learning from imbalanced data sets,* 10(2018), *Cham: Springer*.

[11] Galer, M., Fernandez, A., Barrenechea, E., Bustince, H., Herrera, F. (2012), *A review on ensembles for the class imbalance problem: Bagging-, Boosting-, and Hybrid-based approaches, IEEE Transactions on Systems, Man, and Cybernetics*, Part C, 42, 463-484.

[12] Gicic, A., Donko, D., Subasi, A. (2024), *Time sequence deep learning model for ubiquitous tabular data with unique 3D tensors manipulation, Entropy*, 26, 783.

[13] Han, H., Wang, W., Mao, B. (2005), *Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. In Proceedings of the Conference on Aristophanes Upstairs and Downstairs*, Magdalen Coll, Oxford, UK, 16-18.

[14] He, H., Bai, Y., Garcia, E.A., Li, S. (2008), *ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In Proceedings of the 2008 IEEE International Joint*

Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China, 1-8 June 2008.

[15] He, H., Garcia, E.A. (2009), *Learning from Imbalanced Data. IEEE Transaction on Knowledge and Data Engineering*, 21, 1263-1284.

[16] Jolliffe, I.T. (2002), *Principal Component Analysis, Springer Series in Statistics*. Springer, New York, NY, USA.

[17] Kim, Y., Baik, B., Cho, S. (2016), *Detecting financial misstatements with fraud intention using multi-class cost-sensitive learning. Expert Systems with Applications*, 62, 32-43.

[18] Lee, H.K., Kim, S.B. (2018), *An overlap-sensitive margin classifier for imbalanced and overlapping data. Expert Systems with Applications*, 98, 72-83.

[19] Liang, X., Jiang, A., Li, T., Xue, Y., Wang, G. (2020*), LR-SMOTE-An improved unbalanced data set oversampling based on K-means and SVM. Knowledge-Based Systems*, 196, 105846.

[20] Lin, W.-C., Tsai, C.-F., Hu, Y.-H., Jhang, J.-S. (2017), *Clustering-based undersampling in class-imbalanced data. Information Science,* 409-410, 17-26.

[21] Liu, X.-Y., Wu, J., Zhou, Z.-H. (2009), *Exploratory Undersampling for Class-Imbalance Learning. IEEE Transactions on Systems, Man, and Cybernetics*, Part B, 39, 539-550.

[22] Moro, S., Rita, P., Cortez, P. (2014), *Bank Marketing* [Dataset]. UCI Machine Learning Repository, https://doi.org/10.24432/C5K306.

[23] Muthukumaran, C., Saravanakumar, R., Susila, S.G. (2025). *Mitigating Anomaly-Based DDoS Attack in Heterogeneous IoT Networks with a Federated Learning Model. Computation and Economic Cybernetics Studies and Research*, 59(3), 180-195.

[24] Nguyen, T.-N., Zeno, G., Lars, S.-T. (2010), *Cost-sensitive learning methods for imbalanced data. International Joint Conference on Neural Networks (IJCNN)*, Barcelona, Spain, 18-23 July 2010.

[25] Phua, C., Lee, V., Smith, K., Gayler, R. (2010), *A comprehensice survey of data mining-based fraud detection research*, arXiv:1009.6119.

[26] Prati, R.C, Batista, G., Silva, D.F. (2015), *Class imbalance revisited: a new experimental setup to assess the performance of treatment methods, Knowledge and Information Systems*, 45, 247-270.

[27] Rodgers, J.L., Nicewander, W.A. (1988), *Thirteen ways to look at the correlation coefficient, The American Statistician*, 42(1), 59-66.

[28] Sáez, J.A., Luengo, J., Stefanowski, J., Herrera, F. (2015), *SMOTE–IPF: Addressing the noisy and borderline examples problem in imbalanced classification by a re-sampling method with filtering, Information Science*, 291, 184-203.

[29] van der Maaten, L., Hinton, G. (2008), *Visualizing data using t-SNE, Journal of Machine Learning Research*, 9, 2579-2605.

[30] Verbraken, T., Bravo, C., Weber, R., Baesens, B. (2014), *Development and application of consumer credit scoring models using profit-based classification measures, European Journal of Operational Research*, 238(2), 505-513.

[31] Wahab, F., Khan, I., Sabada, S. (2024), *Credit card default prediction using ML and DL techniques, Internet of Things and Cyber-Physical Systems* 4, 293-306.

[32] Yeh, I. (2009), *Default of Credit Card Clients* [Dataset]. UCI Machine Learning Repository, https://doi.org/10.24432/C55S3H.

[33] Yeh, I-C., Lien, C.-H. (2009), *The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. Expert Systems with Applications*, 36, 2473-2480.