International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

# Enhancing Element Game Classification: Effective Techniques for Handling Imbalanced Classes

**Putri Taqwa Prasetyaningrum[1,2]*** **Purwanto Purwanto[3]** **Adian Fatchur Rochim[4]**

[1]*Doctoral Program of Information Systems, Universitas Diponegoro, Semarang, Indonesia*
[2]*Faculty of Information Technology, Universitas Mercu Buana Yogyakarta, Yogyakarta, Indonesia*
[3] *Department of Chemical Engineering, Faculty of Engineering, Diponegoro University, Semarang, Indonesia*
[4]*Faculty of Engineering, Universitas Diponegoro, Semarang, Indonesia Indonesia*
* Corresponding author's Email: putritaqwa@students.undip.ac.id

**Abstract:** One strategy for enhancing the promotion of marketing content involves the incorporation of gamification techniques. Existing research that examines the utilization of gamification through the classification approach often overlooks the issue of class imbalance within the dataset. Class imbalance occurs when the number of instances in the minority class differs significantly from those in the dominant class. This discrepancy can cause the classification method to perform worse because most classifiers work best when the class distribution in the dataset is balanced. Multiple studies underscore the significance of addressing class imbalances as a pivotal step toward enhancing the performance of classification algorithms. This study aims to demonstrate the impact of the data imbalance challenge and to identify the superior resampling method for integration into the machine learning process. The dataset employed pertains to users of a mobile banking application in Indonesia and encompasses variables such as gamification elements, demographic information, psychological factors, and customer engagement metrics. The chosen resampling technique for this study is SMOTE-Tomek. The various classification methods utilized include linear regression, K-NN, CART, random forest, SVM, stacking ensemble, and XGBoost. Among these methods, the SMOTE-Tomek resampling approach proves most effective when combined with the stacking ensemble classification. This combination yields a remarkable accuracy rate of 98.7% during 10-fold cross-validation, along with an impressive geometric mean score of 0.99. The thorough evaluation findings demonstrated that the approach presented in this study has a promising practical applicability and effectively identifies unbalanced elementary games.

**Keywords:** Elemen game, Classification, Handling imbalance class.

## 1. Introduction

In the realm of financial services, where customer engagement plays a crucial role in driving profitability, the significance of customer relationship management (CRM) is particularly pronounced. This holds especially true for the banking sector [1]. Enhancing customer satisfaction within the realm of mobile banking (m-banking) has consistently attributed great significance to customer relationship management (CRM). Just as in other facets of marketing, CRM remains in a state of constant evolution and refinement. Furthermore, the extensive adoption of mobile banking among internet users has spurred the integration of a customer relationship management system into electronic banking, facilitating the incorporation of m-banking [2]. The introduction of mobile banking has changed the way individuals relate to each other, driving businesses and banks to launch their own mobile banking websites to communicate and interact with their clients directly [3]. An effective method for enhancing promotional efforts within content marketing involves the utilization of gamification. Gamification entails the application of principles and designs from the gaming world to the educational or training context, with the primary goal of infusing it with greater intrigue and amusement for participants. This, in turn, leads to heightened user engagement [4].

In recent years, the popularity of gamification for systems and services has grown significantly, as can be seen with the increasing number of gamification systems [5]. It is estimated that investment in the gamification system will increase dramatically and reach USD 17.76 billion during 2019–2023 [6]. However, even though many efforts have been made in the field of gamification, knowledge about the process of motivating and involving people in gamification systems is still not fully understood [5]. Additionally, gamification is used as a means to educate employees across all types of industries, create customer engagement for brands and businesses, and encourage people to change their behavior. [7]. In Hamari's (2013) interpretation, gamification is characterized as the integration of gaming mechanisms into alternative contexts or pursuits, all in an effort to amplify their allure and enhance customer retention capabilities [8]. Huotari and Hamari (2017) posit that the state of play constitutes a distinctive experiential state shaped by individual perception and guided by inherent motivation [9]. Motivation involving an individual's actions or endeavors driven by internal incentives. This inclination originates from sentiments of accomplishment, contentment, self-governance, proficiency, and interconnectedness [10]. This study explores the correlation between intrinsic motivation and the components of the system, encompassing an array of tools that, when appropriately harnessed, yield substantial (aesthetic) reactions from participants. The system incorporates seven primary elements: points, levels, leaderboards, badges, onboarding, and engagement loops [11].

Gamification can be applied to solve problems following the banking sector. Banking is a prominent field for empirical and methodological research using AI methodology [12]. Prior empirical banking research has primarily focused on factors influencing decision making [13]. In a machine learning approach each participant engages in game- or non-game-based tasks with significant classification accuracy [14]. Furthermore, the convergence of gamification and machine learning manifests through various avenues within both academic and industrial contexts. Numerous endeavors have been undertaken to harness game design components for the enhancement of machine learning workflows. One notable example is the application of gamification principles to the data labeling procedure, incorporating game elements to amplify user engagement and participation in the process [15]. Subsequently, progressions in machine learning (ML) methodologies and their capacity to amplify the capabilities of other technologies, the amalgamation of ML and gamification [16].

Furthermore, the convergence of gamification and machine learning presents multifaceted interactions within academic and industrial realms. Numerous endeavors have aimed to harness elements of game design to enhance the efficiency of machine learning procedures. An illustrative instance of this is the integration of gamification principles into the data labeling process, strategically incorporating gaming elements to enhance user engagement and their active involvement in the process [15]. Next, advances in machine learning (ML) techniques and their potential to enhance other technologies, ML, and gamification combinations [16]. Class imbalance in gamification poses a major challenge to machine learning [17]. Class imbalance occurs when there exists a substantial disparity between the volume of instances in the minority class and the instances within the majority class. This incongruity can lead to suboptimal performance of the classification algorithm, as most algorithms tend to excel under conditions of relatively balanced class distribution within the dataset. Numerous conducted studies underscore the essentiality of addressing imbalanced class scenarios within datasets, highlighting it as a pivotal stride toward enhancing the effectiveness of classification algorithms.

In recent years, many methods have been developed to improve the classification of unbalanced data. Estabrook & Japkowicz [18] have introduced innovative strategies involving the fusion of expert opinions or classifier-generated hybrid methodologies subsequent to the application of various degrees of over-sampling or under-sampling to the data. While this novel hybrid technique has demonstrated efficacy in addressing imbalanced datasets, it is important to acknowledge the presence of other challenges such as overfitting, potential data loss, and the intricacies associated with the resulting model's complexity. Bekkar et. Al. [19] present a comprehensive framework outlining strategies for managing intrusion detection systems (IDS), categorizing these strategies into five primary clusters: sampling techniques, ensemble learning, cost-sensitive learning, feature selection methodologies, and algorithmic adjustments. Dia & Garcia [20] conducted an examination of cutting-edge technologies and contemporary evaluation metrics utilized to assess learning performance within the context of imbalanced learning scenarios. Bekkar et. Al. [19] assembled an all-encompassing survey of strategies for managing imbalanced datasets, focusing on the vantage point of data mining experts. Another facet that data scientists should consider in tackling the challenge of imbalanced datasets pertains to assessment criteria [20-22].

Numerous investigations have been conducted concerning the issue of addressing imbalanced class distribution. Among the various studies, two distinct approaches have emerged: the data-level approach and the algorithmic-level approach. The data-level strategy is typically undertaken during the pre-processing phase, involving alterations or rectifications to mitigate the class distribution bias inherent within the dataset [23]. In the data-level approach, a frequently employed technique involves the utilization of resampling and data synthesis methods. On the other hand, the algorithmic-level approach involves amalgamating multiple algorithms of a similar type into a cohesive unit. Subsequently, diverse scoring techniques, encompassing bagging, stacking, boosting, and voting, are employed for evaluation [24]. Numerous prior investigations propose addressing the challenge of class imbalance through a data-level strategy. In this approach, the class distribution within the dataset can be adjusted to a more balanced state by either augmenting or reducing the number of instances belonging to specific classes.

Below is a compilation of research endeavors focused on addressing class imbalances, with some of these studies proposing a data-level approach as a resolution. Guo's research delved into the impact of handling unbalanced classes within datasets. Within this study, the synthetic minority oversampling technique (SMOTE) was implemented to rectify class imbalances. The outcomes of the tests revealed that the application of SMOTE led to an enhancement in the performance of the utilized classification algorithm [25]. Oversampling techniques are frequently used to balance datasets [26]. An analogous approach has also been employed in studies conducted by [27-29], These three investigations employ the SMOTE oversampling technique to address class imbalances within their datasets. The outcomes of these studies uniformly assert that the utilization of SMOTE yields enhanced performance for the tested classification algorithms. Moreover, two additional studies opted for the under-sampling technique as a means to mitigate class imbalance challenges within their datasets.

SMOTE has demonstrated effectiveness in a number of applications and served as an inspiration for a number of strategies to address the issue of class imbalance [30, 31]. Nonetheless, in certain real-world application settings (often involving challenging datasets), it could produce outcomes that fall short of potential or, in numerous instances, even work against the intended goals [32, 33]. In general, it has three drawbacks: First, it might exacerbate the over-generalization issue by oversampling noisy data, second, it might oversample uninformative samples, and third, it might increase class boundary overlaps. This is due to the fact that SMOTE oversamples every minority sample without restriction and does so blindly [33].

The detection of lithology has been demonstrated to benefit from the use of SMOTE in conjunction with gradient boosting decision trees (GBDT) [34]. For fluid identification and lithology, DNN in conjunction with the MAHAKIL oversampling algorithm which is based on genetic inspiration and Mahalanobis distance has been utilized [35]. Nevertheless, it is important to remember that conventional oversampling methods may decrease the majority class's classification accuracy due to their susceptibility to noise points and border ambiguity [36, 34].

In contrast, as it greatly affects the classifier's overall performance, balancing the data distribution is an essential component of classification. The issue of imbalanced data has been a subject of investigation in numerous previous research endeavors, where different hybrid sampling methods have been employed. Monte carlo mega-trend-diffusion (MCMTD), Hybrid approach operating at both the data and algorithmic levels (HybridDA), synthetic minority oversampling technique & reverse K-nearest neighbours (SMOTE-RkNN), and combined synthetic oversampling & undersampling technique (CSMOUTE) are some of the techniques covered by these methods. In the case of the MCMTD technique, it utilizes the mega-trend diffusion and gaussian fuzzy strategies to generate fresh instances of the minority class and to oversample the data, respectively [37]. This approach is appropriate for handling imbalanced data and focuses on finding solutions to binary classification issues involving imbalanced data. It also creates virtual samples for the minority class[38].

According to reference, HybridDA is a method that combines the methods of SMOTE oversampling, random undersampling (RUS), and SVM optimisation through grid search [39]. This method effectively merges data-level strategies for sample generation and algorithm-level techniques for optimization. It is primarily designed to address the challenge of imbalanced data in binary class scenarios. In contrast, CSMOUTE involves the synthetic generation and elimination of instances from both the minority and majority classes, as detailed in reference [40]. By using both oversampling of the minority class and undersampling of the majority class, CSMOUTE addresses imbalanced data in binary classification using numerical data types. As mentioned in reference [41], MC-CCR, on the other hand, adopts a unique strategy by first cleaning the data and then using oversampling methods. Though MC-CCR works well for numerical data types in imbalanced multi-class classification, its poor stability may render it less useful in real-world imbalanced data classification settings.

When working with multi-class data, imbalanced data difficulties get more complex due to the numerous interactions between classes. Multi-minorities or multi-majorities are involved in the multi-class classification problem. These relationships may be even more complicated in real life. In multi-class contexts, the difficulties arising from imbalanced data classification are compounded by the fact that the classification task becomes more complex with each new class. Because of the intra-class complexity, binary techniques for imbalanced multi-class data have limits. Apart from dealing with unbalanced data, many kinds of data pose serious obstacles that also impact categorization effectiveness. Different approaches and obstacles may be necessary for numerical, picture, and text data, which may have an impact on classification performance [41, 42].

As an alternative option, a multitude of oversampling techniques rooted in the concept of SMOTE have emerged in recent times [43, 44]. One such method is SMOTE-Tomek links, a fusion of SMOTE as an oversampling technique and Tomek links as an undersampling strategy. The distinct advantage of SMOTE-Tomek links lies in its capacity to more efficiently rectify data imbalances compared to standalone SMOTE, consequently enhancing the accuracy of the minority class [45, 43]. Numerous preceding research endeavors have demonstrated the efficacy of the SMOTE-Tomek Links technique in the realm of classification [45, 46].

In the investigation focused on DNA methylation classification [25], the application of the SMOTE-Tomek links technique yielded performance metrics such as recall, precision, and F1 scores surpassing the 90% threshold. Similarly, in the investigation conducted by [45] involving the identification of faults in electric rotary machines, commendable performance metrics exceeding 97% were attained.In the study undertaken by Chandra, the implementation of SMOTE-Tomek links demonstrated its capacity to enhance the efficiency of India's air quality prediction model. Put differently, the method proposed proved effective in addressing the challenges posed by missing values and class imbalance [47].

In the studies conducted by [48, 49] Tomek links (TL) were employed to address class imbalance issues within datasets. The outcomes of employing this under-sampling approach exhibited an enhancement in the geometric mean (gmean) value of the classification algorithm. Furthermore, evaluation metrics such as accuracy, area under the curve (AUC), true negative rate (TNR), and true positive rate (TPR) demonstrated an appreciable increase. Building upon prior research findings, this study intends to employ a hybrid approach of under-sampling and oversampling to effectively tackle the class imbalance challenge inherent in the dataset. The fundamental concept revolves around mitigating the perils of overfitting and potential loss of crucial dataset information. Hence, this study introduces a novel approach that combines the utilization of Tomek links (TL) [50] and synthetic minority over-sampling technique (SMOTE)) [50] to address the challenge of class imbalance present in datasets, with a specific focus on datasets concerning users of mobile banking applications.

This research offers three significant contributions. Primarily, the method introduced by the researcher stands as a potential solution to address class imbalance issues in the classification of game elements, particularly concerning the classification quandaries related to game elements within mobile banking applications. Secondly, the proposed method demonstrates an ability to elevate the performance metrics of the employed classification algorithm, as evidenced in this study. Lastly, this study holds the potential to serve as a foundational reference for future investigations pertaining to the mitigation of class imbalance complexities within datasets, particularly within the scope of data mining research, notably within the banking sector.

## 2. Material and method

### A. MATERIAL

This study conducted a survey among 451 active mobile banking users in various banks across Indonesia. The data was collected from users in yogyakarta, central java, and north sumatra. The survey encompassed a wide range of information, including demographics (gender, age, education, occupation), mobile banking usage (weekly frequency and hours spent online), customer engagement, and psychological factors (self-efficacy, accountability, belongingness). In addition, we assessed the impact of various game elements such as announcements, points, prizes, rankings, badges, scores, tasks, feedback, leaderboards, offers, timers,

Table 1. The dataset samples

| Gender | Education | Occupation | Point | Reward |
|---|---|---|---|---|
| Male | Senior High School | Student | 5 | 3 |
| Female | Master | Private | 5 | 3 |
| Female | Bachelor | Student | 5 | 4 |
| Female | Bachelor | Student | 5 | 4 |
| Male | Bachelor | Student | 5 | 5 |
| Male | Senior High School | Student | 5 | 5 |

levels, social interactive sharing, penalties, avatars, lotteries, virtual prizes, epic meanings, information, and random prizes [51]. Examples of datasets can be seen in Table 1 and Table 2.

In Table 1 is an example dataset that contains information about several data samples. The data consists of several attributes, namely:

a) Gender: Indicates the gender of the individual, which can be either "Male" or "Female".
b) Education: Indicates an individual's education level, such as "Senior high school”, "Masters" or "Bachelor's".
c) Occupation: Indicates the type of individual job, for example "Student" or "Private".
d) Point: Is a number that shows the point value of an activity or achievement.
e) Reward: Is a number that shows the amount of reward given to individuals.

This example dataset consists of six different rows of data. Each row represents one data sample that contains information about a particular individual. This data can be used for further analysis, such as grouping by gender, education, or occupation, as well as looking at the relationship between point values and rewards given.

Table 2 provides descriptions of the datasets for the studies involving the various input features and their corresponding descriptions. The dataset contains nominal and numeric data. This data set relates to user engagement and behavior in systems that use gamification and reward based elements to influence user actions and decisions. Numerical values represent different levels of agreement or disagreement with a particular feature or aspect of the system. Data sets can be used to analyze user preferences, interactions and responses.

Table 2. The dataset description

| Input Feature | Unit | Description |
|---|---|---|
| Gender | Nominal | Male |
| | | Female |
| Age | Nominal | 18 – 24 |
| | | 25 – 34 |
| | | 35 – 44 |
| | | 45 – 54 |
| | | Over 55 years |
| Education | Nominal | Senior High School |
| | | Diploma |
| | | Bachelor |
| | | Master |
| | | Doctoral |
| Occupation | Nominal | Student |
| | | Government employees |
| | | BUMN |
| | | Private |
| | | Self-employed |
| Ability to use mobile banking services (within 1 week) | Nominal | Very low |
| | | Low |
| | | Medium |
| | | Tall |
| | | Very high |
| Hours in 1 week using the internet/social networks/video games and the like | Nominal | 1 – 3 hours |
| | | 3 – 6 hours |
| | | 6 – 9 hours |
| | | 9 – 11 hours |
| | | Above 11 hours |
| User Engaggement, Psychological, Announcements, Point,Reward,Rank,Badge,Score, Task,Feedback, Leaderboard, Hunt for offers, Level, Timer, Share, Social Interactive, Penalty, Avatar, Lottery, Virtual reward, Epic meaning, Informing, Random reward | Numerical | 1=Strongly disagree |
| | | 2=Disagree |
| | | 3=Neutral |
| | | 4=Agree |
| | | 5=Strongly agree |

## B. METHOD

In this research, six different stages were used, as depicted in Fig. 1. These stages include: (1) Data collection; (2) Pre-processing; (3) Missing value imputation; (4) Data balancing; (5) Classification; and (6) Evaluation. A detailed explanation of each stage is presented below.

Received: October 15, 2023. Revised: November 17, 2023. 

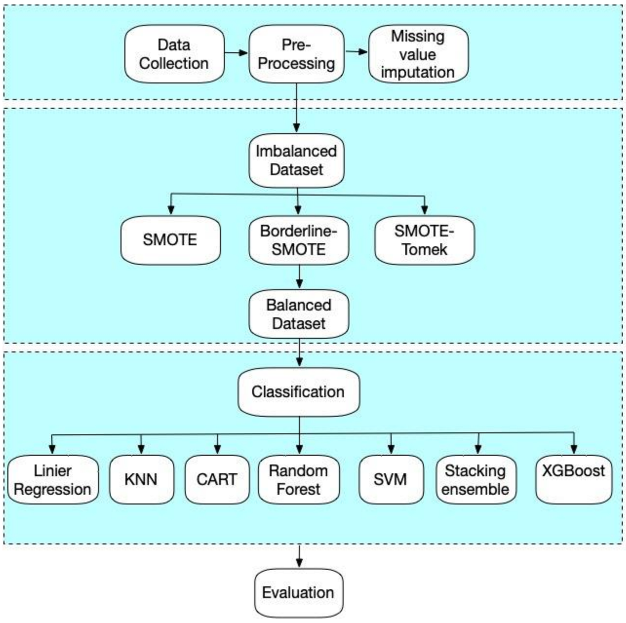


Figure. 1 Research design

### 2.1 Data collection

The first stage of this research is to collect datasets. Collecting data through measurement is very important because it allows the identification of relationships between empirical observations and quantitative mathematical expressions. The data used in this study is primary data obtained through questionnaires distributed to mobile banking users at various banks in Indonesia. The research sample consisted of 451 people, all of whom are active users of mobile banking applications. Data collection was carried out by mobile banking users located in Yogyakarta, Central Java and North Sumatra, Indonesia.

### 2.2 Preprocessing

Preprocessing is a series of processes to ensure the dataset is clean enough to be processed. Raw data sets may contain inconsistent data, noise, incorrect data formats, duplicate or redundant data, or missing values. therefore, this situation can reduce the effectiveness and reliability of data mining results because the dataset is not conditioned to be ready for further processing. To ensure that the data is free from unwanted conditions, it is necessary to carry out the preprocessing stage.

### 2.3 Imbalance data

Several resampling techniques, including a mixture of the two resampling techniques, were used in this study to carry out the resampling procedure.

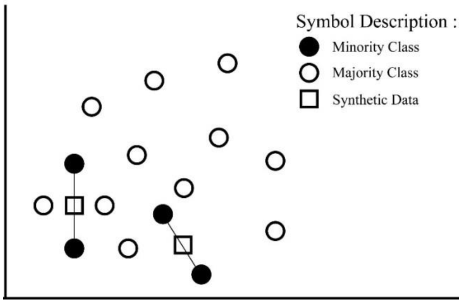


Figure. 2 SMOTE mechanism

SMOTE is the only resampling technique utilized; Tomek links is added to the mix. To increase the number of instances of minority classes, SMOTE is employed. In the majority of classes, noise samples are minimized by the application of Tomek Link and Borderline. SMILESThe main idea behind the SMOTE method is to make a fake version of the desired duplication percentage between the minor data and k randomly chosen nearest neighbors. Then, for each record in the minor class, use that value to figure out which data is next to which. Increasing the quantity of data in the minority class is the aim. An illustration of the SMOTE mechanism can be found in Fig. 2.

1) SMOTE

The SMOTE method's basic idea is to identify the adjacency between data for each record in the minor class, or the k-nearest neighbour value. Then, once the data has been synthetically generated, the desired proportion of duplication between the minor data and the k-nearest neighbours is randomly chosen. The objective is to get more information about the minority class. An illustration of the SMOTE mechanism can be found in Fig. 2.

Class imbalance occurs when the number of instances in the minority class differs significantly from those in the dominant class. This discrepancy can cause the classification method to perform worse because most classifiers work best when the class distribution in the dataset is balanced [52]. SMOTE was first presented by Chawla et al. as a remedy for data imbalance problems [53], Random interpolation is performed between each target class's feature space in the sample and its nearest neighbor, entailing the production of a new sample [30]. This process can augment the count of minority classes, thereby enhancing the classifier's ability to generalize [52] [30]. Many oversampling methods have been developed using SMOTE as a basis [44], including the SMOTE-Tomek link [44, 43]. This method combines SMOTE oversampling and Tomek links

Received: October 15, 2023. Revised: November 17, 2023. 

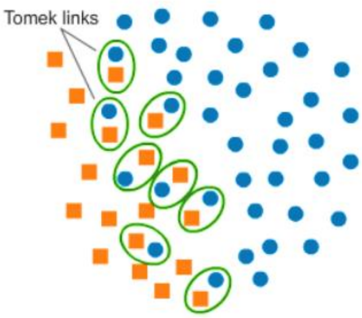


Figure. 3 Tomek links mechanism

undersampling techniques [44]. SMOTE generates artificial data for the minority class, while concurrently, Tomek Links eliminates data marked as Tomek Links from the majority class.

The SMOTE step begins by determining the number of nearest neighbors (k), then calculating the shortest distance between the randomly selected data from the minority class ($x_{ci}$) and the k-nearest neighbors data ($x_{ki}$) using the Euclidean distance in Eq. (1) [52, 54]. Furthermore, based on the shortest distance, synthetic sample data ($x_{si}$) is generated for the minority class using (4) [30]:

$$X_{si} = X_{ci} + r\,(X_{ci} - X_{ki}) \quad (1)$$

The process is stopped when the data for each class is balanced [52, 54].

The first step in Tomek links is to select a pair of samples with the minimum Euclidean distance from the k-nearest neighbors, where each sample comes from a different class ($x_g$, $x_h$). The sample pair $x_g$, $x_h$ is a Tomek Link, if no sample $x_k$ satisfies the following condition with respect to Euclidean distance

$$d(x_g, x_k) < d(x_g, x_h) \quad \text{or} \quad d(x_h, x_k) < d(x_g, x_h) \quad (2)$$

Subsequently, Tomek link samples belonging to the majority class were eliminated from the dataset. The procedure concludes upon achieving class balance [43].

2) Tomek links

The concept behind Tomek links involves removing samples with noise and prioritizing the majority class to achieve a balanced reduction. In essence, mitigating sample noise is crucial to curbing misclassification and maintaining classifier effectiveness. As a result, for this study, the under-sampling technique chosen was Tomek links, as depicted in Fig. 3 below, illustrating its mechanism.

Subsequent to data pre-processing, the next phase involves classification, wherein multiple algorithms are employed for testing. The evaluation of classification algorithms in this research encompasses logistic regression, K-nearest neighbor (K-NN), CART, random forest, support vector machine (SVM), and Stacking. The testing procedure is bifurcated into two distinct scenarios. In the initial scenario, data division entails an 80% training and 20% testing composition. Conversely, the second scenario employs a 10-fold cross-validation approach. Both scenarios serve the primary purpose of assessing the consistency of the classification algorithm's performance. Furthermore, due to the constrained dataset quantity, these two scenarios are employed to ensure a comprehensive evaluation.

The final phase of this study involves the evaluation and juxtaposition of outcomes. This stage encompasses a systematic comparison of classification techniques across several phases. The initial phase involves contrasting classification methods devoid of resampling. Subsequently, the second phase entails a comparison of classification methods integrated with Tomek Links. The third phase encompasses the evaluation of classification methods augmented with SMOTE. Lastly, the fourth phase entails a comprehensive comparison of classification methods leveraging both Tomek and SMOTE links. Within this assessment framework, accuracy and geometric mean (g-mean) serve as pivotal benchmarks. These indicators are chosen due to their comprehensive and pertinent applicability in addressing the complexities of class imbalance scenarios[55]. Classification accuracy is a measure of the degree of confidence in the classification result following an evaluation. This score is calculated by multiplying the true positive rate (TPR) and the true negative rate (TNR), much like the geometric mean (G-mean). A thorough assessment of the overall precision for the majority or minority class can be obtained from the mean G statistic [56]. The following formula illustrates accuracy and the G-mean formula.

### 2.4 Classification

Upon completing the data cleansing, often referred to as the preprocessing phase, the focal phases of the research unfold, encompassing the utilization of distinct classification techniques. The methods employed in this study include Linear Regression, K-NN, CART, Random Forest, SVM, Stacking ensemble, and XGBoost.

The input space is divided into discrete sections by the linear classification approach, which is indicated by linear decision boundaries. Notable examples of linear classification techniques that have acquired significant momentum include logistic regression, separating hyperplanes, multiple linear regression (MLR), and linear discriminant analysis (LDA). This method works especially well when classes can be divided into linear sections [57]. Nonetheless, numerous real-world scenarios present a challenge wherein the data isn't amenable to linear separation. Additionally, instances arise where data points are closely spaced, demanding a highly nonlinear decision boundary for effective data segregation [58].

Linear regression embodies the concept of predicting the conditional mode of the response variable Y, based on a collection of predictors x, by representing it as a linear function of those predictors [59]. The nearest neighbor classification, commonly referred to as K-nearest neighbor (KNN), operates on the premise that the nearest pattern to the target pattern x, for which we seek a label, imparts valuable label insights [19]. The K-NN method is widely used in data mining as one of the simplest methods for classifying.

CART implements numerical separation and Regression can build trees based on CART[60]. The researcher adopts the newly developed model and trains the first CART using the current feature and all features sequentially [61]. Random Forest functions as a classifier composed of an ensemble of tree-structured classifiers, each characterized by independence and the utilization of random vectors. Each individual tree contributes a singular vote toward the prevailing class among the inputs [62]. Support vector machine (SVM) represents a supervised machine learning algorithm suitable for addressing classification and regression tasks. In this method, every data point is graphed as a point within an n-dimensional space (where n denotes the count of feature dimensions), with each feature's value translating to a specific coordinate. The objective of classification entails the identification of a hyperplane that optimally segregates the two classes. SVM is designed to determine a hyperplane that maximizes the separation margin between the nearest data points of one of the classes [63, 64].

Sun's study in 2018 demonstrated the application of ensemble stacking for predicting river ice breakout dates. This approach holds potential for addressing similar challenges in the realm of river ice forecasting [65]. Ensemble stacking learning is structured with two distinct tiers, denoted as the primary base-classifier level and the subsequent meta-classifier level [66, 67]. Within the base-level classifiers, the training dataset is employed for both model training and prediction. In contrast, the meta-classifier employs its meta-data for training, and it maps the outcomes from the base-level classifier to the corresponding actual classifier labels. A similar technique was utilized by the Netflix team, known as "The ensemble," where the top-performing team's submission, determined by its accuracy, was integrated [68, 69].

XGBoost exhibits exceptional predictive capabilities due to its foundation in a gradient-boosted decision tree ensemble algorithm. Chen and Guestrin (2016) highlight its prowess in managing sparse data, optimizing memory usage through problem approximation, extending tree fitting to residuals, and achieving enhanced scalability [70]. As an additional motivation for its selection, XGBoost has been shown to outperform other machine learning techniques, such as traditional classification and regression trees, both in road safety[71] and in other fields [72], however to date there have been relatively few road safety applications for road safety related tasks compared to more other machine learning algorithms.

### 2.5 Evaluation

The evaluation and comparison process is the final step of this study project. The categorization techniques will be evaluated against one another in multiple phases during this phase. The first is a comparison of classification methods that do not use resampling; the second is a comparison of classification methods using Tomek links; the third is a comparison of classification methods using SMOTE; and the fourth is a comparison of classification methods using Tomek and SMOTE links. Since accuracy and geometric mean (g-mean) are most comprehensive when considered in the context of class imbalance, they are utilized as assessment indicators in this study [55]. Classification accuracy is a score that represents the amount of certainty of a data record that has been classified following an evaluation. Like the G-mean, the score is determined by multiplying the true positive rate (TPR) and the true negative rate (TNR). The average G statistic represents the overall accuracy for the minority or majority class [56]. Accuracy and the G-mean formula can be seen in the equation below.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

Table 3. Sample of preprocessing data results

| JK | EDU | PK | PO | RW |
|---|---|---|---|---|
| 1 | 5 | 5 | 5 | 3 |
| 0 | 4 | 4 | 5 | 3 |
| 0 | 5 | 5 | 5 | 4 |
| 0 | 2 | 5 | 5 | 4 |
| 1 | 3 | 5 | 5 | 5 |
| 1 | 3 | 3 | 5 | 5 |

$$G - Mean = \sqrt{TPR \times TNR} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7)$$

Accuracy is used to calculate the accuracy of the classification model [55]. Other metrics that are employed include F1-score, recall, and precision. While recall determines the percentage of inputs that the system correctly recognizes as correct, precision concentrates on the ratio of inputs that the system accurately identifies [73]. The F-1 score, on the other hand, is a weighted average of accuracy and recall.

## 3. Result and discussion

This study explores the impact of data imbalance on classification outcomes and examines resampling techniques as a potential solution. Additionally, diverse evaluation metrics for classification are employed to offer insights into the effects of the evaluation approach. These findings help identify the optimal resampling and classification methods for all models. The implemented models are developed using Python. To assess the classifier's performance on unbalanced data, the initial step involves running the classifier on the original dataset. Subsequently, the data undergoes various oversampling methods to address imbalance before being subjected to classification.

### 3.1 Preprocessing results

Following are the results of data preprocessing that has been done, can be seen in Table 3.

Table 3 shows the results of the data pre-processing process with several columns containing processed values. Here is a description of each column:

a) JK (Gender): number 1 indicates male gender, number 0 indicates female gender.

b) EDU (Education): SMA Denotes the level of education SMA (high school) with a 5, Bachelor denotes a level of education Bachelor (S1 or equivalent) with a number 4. Master denotes a level of education in a masters.

c) Occupation: A number or category that indicates the type of work or employment status of the individual concerned.

d) Point: A numeric variable that describes the user's view of a particular point. Value 1 (Strongly disagree), 2 (Disagree), 3 (Neutral), 4 (Agree), 5 (Strongly agree).

e) Reward: A numeric variable that describes the user's view of a particular reward. Possible values: 1 (Strongly disagree), 2 (Disagree), 3 (Neutral), 4 (Agree), 5 (Strongly agree).

Given the data provided, this table appears to contain information about a group of individuals who share certain attributes, such as gender, education level, occupation, number of points accumulated, and rewards received. This table may be used for further analysis, such as analyzing the relationship between gender, education level, and occupation and performance or participation in the system or application.

### 3.2 Handling imbalance data

Addressing imbalanced data is a significant and prevalent hurdle in both data analysis and machine learning endeavors. This challenge arises when the distribution of classes or target variables within a dataset is heavily skewed, resulting in the underrepresentation of one or more classes and the overrepresentation of others. This imbalance introduces the potential for biased models, limited generalizability, and unreliable predictions, particularly concerning the minority class.

Fig. 4 illustrates that the Y axis shows the number of respondents, and the X axis shows the game elements. Fig. 4 shows this uneven distribution, indicating a fairly large data imbalance between game elements. The limited occurrence of certain elements may have a negligible impact on the decision-making process, whereas more general elements will have a significant influence on the dynamics of the game. Therefore, the main thing to do now is to solve this data imbalance problem. Additionally, the image above displays a catalog of game elements selected by mobile banking app users, each accompanied by its own frequency count. These game elements serve as integral components used in game design and development to strengthen player

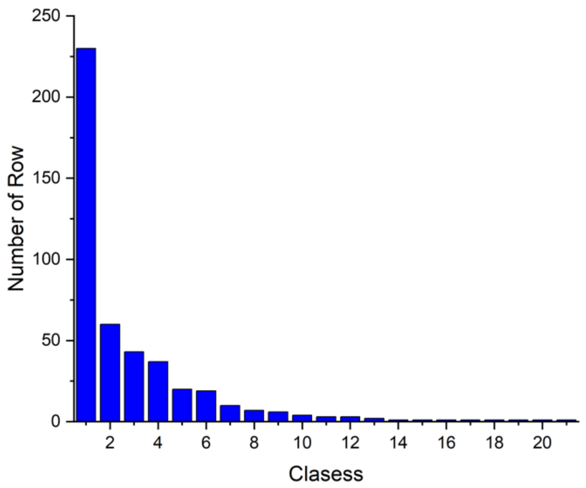


Figure. 4 Class distribution of the dataset

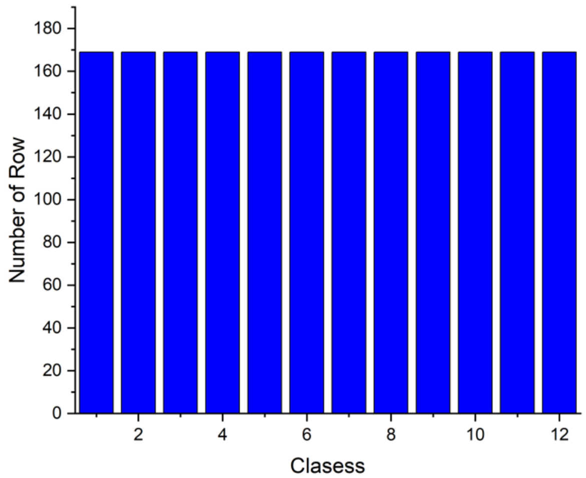


Figure. 5 After handling imbalance data

engagement, interaction, and overall experience. A higher count indicates a more frequent occurrence of the related element in the game.

Several methodologies and techniques exist to tackle the quandary of data imbalance. In this phase, the Smote-Tomek method is executed utilizing an oversampling technique. Smote-Tomek is employed to establish equilibrium between class frequencies by augmenting data in the minority class. The class with the scarcest training data assumes this role.

Fig. 5 depicts the outcomes of employing the SMOTE (Synthetic minority over-sampling technique) algorithm in conjunction with simplex sampling on a 10-class dataset. SMOTE serves as an oversampling method designed to counteract class imbalance within datasets. This notation signifies the usage of SMOTE with the simplex sampling technique, applied to a dataset characterized by two dimensions (features). The visual presentation in Fig. 5 illustrates the instance count for each class in the dataset post the implementation of SMOTE using simplex sampling. Notably, all classes (ranging from 0 to 9) now exhibit 169 instances each. As an example, Class 0 in the dataset showcases a total of 169 instances, which constitutes a relative proportion of 10,000% within the resampled dataset (given the presence of 169 instances for each class out of a cumulative 1690 instances).

This pivotal step endeavors to rectify the issue of disparate data distribution through the amalgamation of oversampling (SMOTE) and undersampling (Tomek links) techniques, culminating in the creation of a more harmonized dataset suitable for classification purposes. The overarching goal revolves around heightening the performance and precision of machine learning models when grappling with datasets characterized by imbalanced class distributions.

### 3.3 Original dataset classification results

The classification techniques employed in this study encompass linear regression, K-NN, CART, random forest, SVM, stacking ensemble, and XGBoost. Subsequently, the classification outcomes were subjected to assessment via two distinct scenarios: firstly, a 10-fold cross-validation was executed, producing an average accuracy score derived from the iterations within the k-fold mechanism. Secondly, data partitioning ensued in an 80 – 20 ratio, yielding several metrics like Precision, Recall, and F1 Scores.

Table 4 provides an overview of the classification results for the different classifier techniques and algorithms that were used. Remarkably, every classifier technique or algorithm demonstrates a different level of expertise in categorization tasks. The classification results are expressed as an accuracy percentage. Table 4 is a thorough source of information on how well the model performs with data that exhibits class imbalance.

In Table 4, the linear regression method attains an accuracy of 65.08%. The K-nearest neighbors (KNN) method achieves an accuracy of 68.25%. The classification and regression trees (CART) method secures an accuracy of 69.05%. The random forest method exhibits the highest accuracy, reaching 72.22%. The support vector machine (SVM) method garners an accuracy of 61.90%. The stacking ensemble method records an accuracy of 69.10%. Lastly, the XGBoost method attains an accuracy of 66.15%. Despite the dataset's imbalanced nature, all classifier scores surpass the 50% threshold, indicating a favorable outcome. However, this

Table 4. Classification result's on original dataset

| Classifier | Imbalanced Dataset | | | | | | |
|---|---|---|---|---|---|---|---|
| | ***Data Splitting*** | | | | | | ***10-fold cross validation*** |
| | Precision | Recall | Geometric Mean | F1 | | | Average Accuracy |
| | | | | Class0 | Class1 | Class2 | |
| Linier Regression | 65% | 65 % | 0.69 | 60 % | 70 % | 65 % | 65.08 % |
| K-NN | 69% | 69 % | 0.73 | 62 % | 71 % | 69 % | 68.25 % |
| CART | 70% | 70 % | 0.75 | 69 % | 73 % | 66 % | 69.05 % |
| Random Forest | 79% | 77 % | 0.76 | 75 % | 82 % | 75 % | 72.22 % |
| SVM | 66% | 66 % | 0.73 | 58 % | 74 % | 63 % | 61.90 % |
| Stacking ensemble | 70% | 75 % | 0.80 | 59 % | 75 % | 64 % | 69.10 % |
| XGBoost | 67% | 65 % | 0.79 | 66 % | 73 % | 66 % | 66.15 % |

achievement isn't adequate to conclude that these are strong classification results, and there remains potential room for refinement.

### 3.4 Classification with resampling method

As was already mentioned, there is an imbalance in the class distribution, which suggests that each class is not represented proportionately. The F1 value makes the significance of this mismatch on classification clear. The harmonic mean of precision and recall is represented by the F1 score, which offers important information about classifier performance within each class. As an example, K-NN shows a comparatively lower F1 score for Class 0 while the other two classes receive scores higher than 50%.

This study revolves around datasets involving multi-classification challenges. The deleterious impact of uneven class distribution becomes apparent through diminished classifier performance, as underscored by the outcomes presented in Table 4, which reveal lower accuracy scores. Consequently, the rectification of imbalanced data emerges as a requisite for achieving more favorable classification outcomes. Table 5 and Table 6 detail the outcomes of

Table 5. SMOTE resample method's result

| Classifier | SMOTE | | | | |
|---|---|---|---|---|---|
| | ***Data Splitting*** | | | | ***10-fold cross validation*** |
| | Recall | Specificity | Geometric Mean | F1 Score | |
| Linier Regression | 97% | 100% | 0.98 | 97 % | 93.9% |
| K-NN | 96% | 100% | 0.98 | 96 % | 94.5% |
| CART | 94% | 99% | 0.97 | 94 % | 93.1% |
| Random Forest | 97% | 100% | 0.98 | 97 % | 95.6% |
| SVM | 97% | 100% | 0.98 | 97 % | 94.3% |
| **Stacking ensemble** | **99 %** | **100%** | **0.99** | **98 %** | **95.7%** |
| XGBoost | 96% | 100% | 0.98 | 96 % | 92.8% |

individual machine learning techniques when applied to balanced data, which underwent resampling using diverse methodologies.

Table 5 displays the outcomes obtained from the SMOTE resampling method applied across various classifier types, along with the performance evaluation metrics. The SMOTE resampling technique was implemented, and the performance evaluation employed data splitting in conjunction with 10-fold cross-validation.

The ensuing descriptions encapsulate the performance evaluation outcomes for each classifier:

a) Linear Regression exhibits outstanding results, characterized by high accuracy levels and consistent capability in recognizing positive classes.
b) K-NN also demonstrates commendable performance, showcasing high accuracy rates and adeptness in identifying positive classes.
c) CART delivers favorable outcomes, marked by high accuracy rates and proficient identification of positive classes.
d) Random Forest showcases exceptional performance, with elevated accuracy rates and a consistent ability to identify positive classes.
e) SVM also yields promising results, featuring high accuracy levels and adeptness in identifying positive classes.

Table 6. Borderline-SMOTE resampling method's result

| Classifier | SMOTE | | | | |
|---|---|---|---|---|---|
| | *Data Splitting* | | | | ***10-fold cross validation*** |
| | Recall | Specificity | Geometric Mean | F1 Score | |
| Linier Regression | 96% | 100% | 0.98 | 96 % | 93.9% |
| K-NN | 95% | 100% | 0.98 | 95 % | 94.5% |
| CART | 94% | 99% | 0.97 | 93 % | 90.8% |
| Random Forest | 97% | 100% | 0.98 | 97 % | 94.6% |
| SVM | 95% | 100% | 0.98 | 96 % | 92.7% |
| **Stacking ensemble** | **98 %** | **100%** | **0.99** | **98 %** | **95.8%** |
| XGBoost | 94% | 100% | 0.98 | 95 % | 93.6% |

f) Stacking ensemble demonstrates superior performance, boasting high accuracy rates and a consistent proficiency in recognizing positive classes.
g) XGBoost produces positive results, characterized by high accuracy levels and capable identification of positive classes.

The SMOTE resampling method has proven successful in elevating classification performance across all evaluated classifier types. This is evident through the heightened accuracy rates and improved capability in recognizing positive classes.

Table 6 exhibits the assessment outcomes stemming from the utilization of the Borderline-SMOTE resampling technique across diverse classifications, employing the 10-fold cross-validation methodology. This table encompasses a range of evaluation metrics, encompassing recall, specificity, geometric mean, and F1 score, corresponding to each classification. The table facilitates a rapid juxtaposition of the classifier's performance when integrated with the Borderline-SMOTE resampling technique, illustrating their efficacy in mitigating class imbalance and accurately classifying instances within a given dataset. The evaluation metrics presented within this table hold profound significance in appraising the efficacy of the Borderline-SMOTE resampling technique in enhancing classification performance and enabling well-informed decisions pertaining to model selection.

Table 7. SMOTE-TOMEK resampling method's result

| Classifier | SMOTE | | | | |
|---|---|---|---|---|---|
| | *Data Splitting* | | | | ***10-fold cross validation*** |
| | Recall | Specificity | Geometric Mean | F1 Score | |
| Linier Regression | 97% | 100% | 0.98 | 97 % | 97.1% |
| K-NN | 96% | 100% | 0.98 | 96 % | 95.2% |
| CART | 95% | 99% | 0.97 | 94 % | 94.0% |
| Random Forest | 98% | 100% | 0.99 | 98 % | 96.7% |
| SVM | 97% | 100% | 0.98 | 97 % | 95.8% |
| **Stacking ensemble** | **99 %** | **100%** | **0.99** | **99 %** | **98.7%** |
| XGBoost | 98% | 100% | 0.99 | 98 % | 96.8% |

Table 7 showcases the outcomes resulting from the application of the SMOTE-TOMEK resampling technique across various classifier types, accompanied by metrics for evaluating classification performance. The resampling process involved the implementation of the SMOTE-TOMEK method, while performance evaluation was conducted through data partitioning utilizing 10-fold cross-validation. The subsequent narrative elucidates the performance assessment outcomes for each classifier:

a) Linear regression yields exceptional outcomes, characterized by a high level of accuracy and a consistent proficiency in identifying positive classes.
b) K-NN also exhibits commendable performance, boasting a high degree of accuracy and adeptness in recognizing positive classes.
c) CART delivers positive results, featuring elevated accuracy levels and proficient identification of positive classes.
d) Random forest demonstrates outstanding performance, marked by high accuracy levels and consistent capability to recognize positive classes.
e) SVM also presents positive results, with a high degree of accuracy and proficient identification of positive classes.

Received: October 15, 2023. Revised: November 17, 2023. 

Table 8. Resampling method comparasion

| **Resampling method + Classifier** | ***Data Splitting*** | | ***10-fold cross validation*** |
|---|---|---|---|
| | Geometric Mean | F1 Score | Average Accuracy |
| SMOTE+ Stacking ensemble | 0.99 | 98% | 95.7% |
| Borderline SMOTE+ Stacking ensemble | 0.99 | 98% | 95.8% |
| SMOTE-Tomek + Stacking ensemble | **0.99** | **99%** | **98.7%** |

f) Stacking ensemble showcases exemplary performance, characterized by a notable accuracy level and consistent proficiency in recognizing positive classes.
g) XGBoost achieves favorable results, showcasing a high accuracy level and proficient ability to identify positive classes.

Collectively, the implementation of the SMOTE-TOMEK resampling technique has effectively enhanced the classification performance across all classifier types scrutinized. This enhancement is underscored by elevated accuracy levels and improved proficiency in recognizing positive classes. Notably, the stacking ensemble stands out with outstanding performance, boasting a high degree of accuracy and an exceptional aptitude for identifying positive classes.

### 3.5 Discussion

The results reveal that each approach exerts a substantial influence on classification performance, whether achieved via oversampling, undersampling, or hybrid sampling techniques. These outcomes are presented in Table 8.

Table 8 presents a comparative overview of the employed resampling methods, specifically SMOTE and SMOTE-Tomek, along with performance evaluation metrics encompassing geometric mean, F1 score, and average accuracy. The assessment of performance is executed through data splitting utilizing the 10-fold cross-validation methodology. Below is a comprehensive analysis detailing the performance comparison across each resampling method:

a) SMOTE + stacking ensemble
The amalgamation of the SMOTE resampling method and Stacking Ensemble yields exceptional outcomes, characterized by a high accuracy level and adeptness in consistently identifying positive classes.
b) Borderline SMOTE + stacking ensemble
With the Borderline SMOTE resampling method, Stacking Ensemble produces commendable results.
c) SMOTE-Tomek + stacking ensemble
Upon integration with the SMOTE-Tomek resampling method, Stacking Ensemble manifests outstanding results, marked by a high accuracy level and exceptional proficiency in recognizing positive classes.

All three resampling techniques SMOTE, Borderline SMOTE, and SMOTE-Tomek perform admirably in terms of improving classification accuracy. However, when combined with the stacking ensemble method, SMOTE-Tomek + stacking ensemble produces somewhat better outcomes than SMOTE alone. An improved average accuracy and F1 score attest to this difference. The combination of SMOTE-Tomek + stacking ensemble is therefore found to be more effective in improving classification performance.

## 4. Conclusion

The aim of this research is to address the problem of data imbalance in classification tasks by introducing a new and efficient pre-processing technique. These results suggest that combining SMOTE-TOMEK with stacking ensembles may be a good way to handle this particular dataset. This methodology was created to improve the handling of class imbalances in multi-class datasets, which is a major challenge in the field of machine learning. The effectiveness of the proposed hybrid sampling solution, demonstrated through the curation of user data from a mobile banking application, which includes demographics and gaming elements, provides strong evidence that this approach significantly improves the performance of multi-class imbalanced data classification tasks in the banking sector in particular.

In particular, this study shows that the proposed methodology goes beyond conventional approaches to dealing with imbalanced data. The proposed approach outperforms other standard methods, as it shows its effectiveness in managing imbalances in multiclass data classification. The results of this investigation illustrate the efficiency of pre-processing techniques in overcoming the challenges associated with the classification of imbalanced multiclass text data. Through the integration of oversampling and undersampling methods, a more robust resolution for class imbalance in multiclass datasets can be provided. This technique is easy to apply to a variety of text data classification tasks and promises to increase the precision and reliability of machine learning or deep learning algorithms.

On the other hand, it is important to note that this research could be further expanded by conducting larger experiments that include ensemble classification and adding additional resampling methods. The addition of datasets about mobile banking apps promises to deepen insights and improve classification performance by allowing users to customize features more. Future research should focus on testing full performance on various data sets to learn more about how this approach can be used in other situations and to discover any problems that may arise.

Despite the limitations of this study, it has provided a foundation for future investigations regarding handling class imbalances in multiclass datasets. The authors recommend more research in this area since the suggested hybrid sampling approach has the potential to enhance the performance of deep learning or machine learning algorithms on a variety of unbalanced datasets.

## Conflicts of interest

The authors declare no conflict of interest.

## Author contributions

Conceptualization, PTP and AFR; methodology, PTP and PP; software, PTP; validation, AFR. and PP; formal analysis, PTP; investigation, PTP and P; resources, PTP; data curation, PTP; writing original draft preparation, PTP; writing review and editing, PTP; visualization, PTP; supervision, PP. and AFR; project administration, PTP.

## Acknowledgments

The author would like to thank the Doctoral Program in Information Systems, Diponegoro University, Semarang and the Faculty of Information Technology, Mercu Buana University, Yogyakarta, for all their support for our research. Lastly, we would like to express our gratitude to the editors, reviewers, and publishing journal personnel for their thoughtful consideration and direction during the peer review and publication processes. Their professionalism and knowledge have been crucial to the completion of this project.

## References

[1] G. K. Agrawal and D. Berg, “The Development of Services in Customer Relationship Management (CRM) Environment from ‘Technology’ Perspective”, *J. Serv. Sci. Manag.*, Vol. 02, No. 04, pp. 432–438, 2009, doi: 10.4236/jssm.2009.24052.

[2] H. Hamidi and M. Safareeyeh, “A model to analyze the effect of mobile banking adoption on customer interaction and satisfaction: A case study of m-banking in Iran”, *Telemat. Informatics*, Vol. 38, No. June 2018, pp. 166–181, 2019, doi: 10.1016/j.tele.2018.09.008.

[3] M. N. Hajli, J. Sims, M. Featherman, and P. E. D. Love, “Credibility of information in online communities”, *J. Strateg. Mark.*, Vol. 23, No. 3, pp. 238–253, 2015, doi: 10.1080/0965254X.2014.920904.

[4] A. Pratomo, “Pengaruh Konsep Gamifikasi Terhadap Tingkat Engagement”, *J. Tour. Hosp. Essentials J.*, Vol. 8, No. 2, p. 63, 2018, doi: 10.17509/thej.v8i2.13740.

[5] J. Koivisto and J. Hamari, “The rise of motivational information systems: A review of gamification research”, *Int. J. Inf. Manage.*, Vol. 45, No. October 2018, pp. 191–210, 2019, doi: 10.1016/j.ijinfomgt.2018.10.013.

[6] Technavio, “Gamification Market by Application, End-users, Need, and Geography - Global Forecast and Analysis 2019-2023. Technavio Report”, 2019.

[7] N. V. W. Paderborn, A. Gustafsson, Norwegian, J. Hamari, P. Parvinen, and A. Haff, “The Great Game of Business: Advancing Knowledge on Gamification in Business Contexts”, *J. Bus. Res.*, Vol. 53, No. 9, pp. 1–16, 2019, doi: DOI: 10.1016/j.jbusres.2019.10.062.

[8] J. Hamari, “Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service”, *Electron. Commer. Res. Appl.*, Vol. 12, No. 4, pp. 236–245, 2013, doi: 10.1016/j.elerap.2013.01.004.

[9] K. Huotari and J. Hamari, “A definition for gamification: anchoring gamification in the service marketing literature”, *Electron. Mark.*, Vol. 27, No. 1, pp. 21–31, 2017, doi: 10.1007/s12525-015-0212-z.

[10] E. L. Deci, “Intrinsic motivation”, *New York and London*, 1975, doi: 10.1057/9781403990037_26.

[11] G. Zichermann and C. Cunningham, “Implementing Game Mechanics Web and Mobile Apps”, 2011. [Online]. Available: http://storage.libre.life/Gamification_by_Desig

Received: October 15, 2023. Revised: November 17, 2023. 

n.pdf

[12] M. Doumposa, Zopounidisab, C. Gounopoulosc, E. Platanakisc, and W. Zhangc, "Operational research and artificial intelligence methods in banking", *Eur. J. Oper. Res.*, pp. 1–16, 2022.

[13] K. Rajaratnam, P. Beling, and G. Overstreet, "Regulatory capital decisions in the Context of consumer loan portfolios ga", *J. Oper. Res. Soc.*, Vol. 68, No. 7, pp. 847–858, 2017, doi: 10.1057/jors.2016.7.

[14] M. Ninaus *et al.*, "Increased emotional engagement in game-based learning – A machine learning approach on facial emotion detection data", *Comput. Educ.*, Vol. 142, No. July, p. 103641, 2019, doi: 10.1016/j.compedu.2019.103641.

[15] A. L'Heureux, K. Grolinger, W. A. Higashino, and M. A. M. Capretz, "A gamification framework for sensor data analytics", In: *Proc. of 2017 IEEE international congress on internet of things (ICIOT)*, pp. 74–81, 2017.

[16] A. Khakpour and R. C. Palacios, "Convergence of Gamification and Machine Learning", *A Systematic Literature Review*, Springer Netherlands, Vol. 26, No. 3, 2021, doi: 10.1007/s10758-020-09456-4.

[17] A. K. Baughman, W. Chuang, K. R. Dixon, Z. Benz, and J. Basilico, "Deepqa jeopardy! gamification: a machine-learning perspective", *IEEE Trans. Comput. Intell. AI games*, Vol. 6, No. 1, pp. 55–66, 2013.

[18] A. Estabrooks and N. Japkowicz, "A mixture-of-experts framework for learning from imbalanced data sets", in *International Symposium on Intelligent Data Analysis*, pp. 34–43, 2001.

[19] M. Bekkar and T. A. Alitouche, "Imbalanced data learning approaches review", *Int. J. Data Min. Knowl. Manag. Process*, Vol. 3, No. 4, p. 15, 2013.

[20] H. He and E. A. Garcia, "Learning from imbalanced data", *IEEE Trans. Knowl. Data Eng.*, Vol. 21, No. 9, pp. 1263–1284, 2009.

[21] G. M. Weiss, "Foundations of imbalanced learning", *Imbalanced Learn. Found. Algorithms, Appl.*, pp. 13–41, 2013.

[22] M. Bekkar, H. K. Djemaa, and T. A. Alitouche, "Evaluation Measures for Models Assessment over Imbalanced Data Sets", *J. Inf. Eng. Appl.*, Vol. 3, No. 10, pp. 27–38, 2013, [Online]. Available: http://www.iiste.org/Journals/index.php/JIEA/article/view/7633

[23] F. Shakeel, A. S. Sabhitha, and S. Sharma, "Exploratory review on class imbalance problem: An overview", *2017 8th Int. Conf. Comput. Commun. Netw. Technol.*, pp. 1–8, 2017.

[24] V. S. Spelmen and R. Porkodi, "A Review on Handling Imbalanced Data", In: *Proc. 2018 Int. Conf. Curr. Trends Towar. Converging Technol. ICCTCT 2018*, No. December, pp. 1–11, 2018, doi: 10.1109/ICCTCT.2018.8551020.

[25] H. Guo, J. Zhou, and C. A. Wu, "Imbalanced learning based on data-partition and SMOTE", *Inf.*, Vol. 9, No. 9, 2018, doi: 10.3390/info9090238.

[26] A. Fernández, S. García, F. Herrera, and N. V. Chawla, "SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary", *J. Artif. Intell. Res.*, Vol. 61, pp. 863–905, 2018, doi: 10.1613/jair.1.11192.

[27] M. S. Islam, M. Arifuzzaman, and M. S. Islam, "SMOTE Approach for Predicting the Success of Bank Telemarketing", In: *Proc. of 2019 4th Technology Innovation Management and Engineering Science International Conference (TIMES-iCON)*, pp. 1–5, 2019, doi: 10.1109/TIMES-iCON47539.2019.9024630.

[28] D. Bajer, B. Zonć, M. Dudjak, and G. Martinović, "Performance Analysis of SMOTE-based Oversampling Techniques When Dealing with Data Imbalance", In: *Proc. of 2019 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 265–271, 2019, doi: 10.1109/IWSSIP.2019.8787306.

[29] E. Kurniawan, F. Nhita, A. Aditsania, and D. Saepudin, "C5.0 algorithm and synthetic minority oversampling technique (SMOTE) for rainfall forecasting in bandung regency", In: *Proc. of 2019 7th Int. Conf. Inf. Commun. Technol. ICoICT 2019*, No. July, pp. 1–5, 2019, doi: 10.1109/ICoICT.2019.8835324.

[30] A. Fernandez, S. Garcia, and F. Herrera, "SMOTE for Learning from Imbalanced Data_ Progress and Challenges, Marking the 15-year Anniversary", *J. Artificial Intell. Res.*, Vol. 61, pp. 863–905, 2018.

[31] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, *Learning from Imbalanced Data Sets*. 2018. doi: 10.1007/978-3-319-98074-4.

[32] J. A. Sáez, J. Luengo, J. Stefanowski, and F. Herrera, "SMOTE-IPF: Addressing the noisy and borderline examples problem in imbalanced classification by a re-sampling method with filtering", *Inf. Sci. (Ny).*, Vol. 291, No. C, pp. 184–203, 2015, doi: 10.1016/j.ins.2014.08.051.

[33] S. Barua, M. M. Islam, X. Yao, and K. Murase, "MWMOTE - Majority weighted minority

oversampling technique for imbalanced data set learning", *IEEE Trans. Knowl. Data Eng.*, Vol. 26, No. 2, pp. 405–425, 2014, doi: 10.1109/TKDE.2012.232.

[34] K. Zhou, J. Zhang, Y. Ren, Z. Huang, and L. Zhao, "A gradient boosting decision tree algorithm combining synthetic minority oversampling technique for lithology identification", *Geophysics*, Vol. 85, No. 4, pp. WA147–WA158, 2020, doi: 10.1190/geo2019-0429.1.

[35] M. He, H. Gu, and H. Wan, "Log interpretation for lithology and fluid identification using deep neural network combined with MAHAKIL in a tight sandstone reservoir", *J. Pet. Sci. Eng.*, Vol. 194, No. May, p. 107498, 2020, doi: 10.1016/j.petrol.2020.107498.

[36] M. Denil and T. Trappenberg, "Overlap versus imbalance", *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, LNAI, Vol. 6085, pp. 220–231, 2010, doi: 10.1007/978-3-642-13059-5_22.

[37] X. Yu, Y. He, Y. Xu, and Q. Zhu, "A Mega-Trend-Diffusion and Monte Carlo based virtual sample generation method for small sample size problem", *J. Phys. Conf. Ser.*, Vol. 1325, No. 1, 2019, doi: 10.1088/1742-6596/1325/1/012079.

[38] A. Zhang, H. Yu, Z. Huan, X. Yang, S. Zheng, and S. Gao, "SMOTE-RkNN: A hybrid re-sampling method based on SMOTE and reverse k-nearest neighbors", *Inf. Sci. (Ny).*, Vol. 595, pp. 70–88, 2022, doi: 10.1016/j.ins.2022.02.038.

[39] M. M. A. Rifaie and H. A. Alhakbani, "Handling class imbalance in direct marketing dataset using a hybrid data and algorithmic level solutions", In: *Proc. of 2016 SAI Comput. Conf. SAI 2016*, pp. 446–451, 2016, doi: 10.1109/SAI.2016.7556019.

[40] M. Koziarski, "CSMOUTE: Combined Synthetic Oversampling and Undersampling Technique for Imbalanced Data Classification", In: *Proc. of Int. Jt. Conf. Neural Networks*, Vol. 2021-July, 2021, doi: 10.1109/IJCNN52387.2021.9533415.

[41] M. Koziarski, M. Woźniak, and B. Krawczyk, "Combined Cleaning and Resampling algorithm for multi-class imbalanced data with label noise", *Knowledge-Based Syst.*, Vol. 204, p. 106223, 2020, doi: 10.1016/j.knosys.2020.106223.

[42] B. Krawczyk, M. Koziarski, and M. Wozniak, "Radial-based oversampling for multiclass imbalanced data classification", *IEEE Trans. Neural Networks Learn. Syst.*, Vol. 31, No. 8, pp. 2818–2831, 2020, doi: 10.1109/TNNLS.2019.2913673.

[43] L. A. Jun and Z. Peng, "Research on Unbalanced Data Processing Algorithm Base Tomeklinks-Smote", In: *Proc. of the 2020 3rd International Conference on Artificial Intelligence and Pattern Recognition*, 2020, pp. 13–17. doi: 10.1145/3430199.3430222.

[44] M. Zeng, B. Zou, F. Wei, X. Liu, and L. Wang, "Effective prediction of three common diseases by combining SMOTE with Tomek links technique for imbalanced medical data", In: *Proc. of 2016 IEEE Int. Conf. Online Anal. Comput. Sci. ICOACS 2016*, No. May 2016, pp. 225–228, 2016, doi: 10.1109/ICOACS.2016.7563084.

[45] E. F. Swana, W. Doorsamy, and P. Bokoro, "Tomek Link and SMOTE Approaches for Machine Fault Classification with an Imbalanced Dataset", *Sensors*, Vol. 22, No. 9, 2022, doi: 10.3390/s22093246.

[46] C. Liu, J. Wu, Y. Song, and W. Hou, "Classifying dna methylation imbalance data in cancer risk prediction using smote and tomek link methods", In: *Proc. of Data Science: 4th International Conference of Pioneering Computer Scientists, Engineers and Educators*, Zhengzhou, China, pp. 1–9, 2018. doi: 10.1007/978-981-13-2206-8_1.

[47] W. Chandra, B. Suprihatin, and Y. Resti, "Median-KNN Regressor-SMOTE-Tomek Links for Handling Missing and Imbalanced Data in Air Quality Prediction", *Symmetry (Basel).*, Vol. 15, No. 4, 2023, doi: 10.3390/sym15040887.

[48] E. AT, A. M, A. M. F, and S. M, "Classification of Imbalance Data using Tomek Link (T-Link) Combined with Random Under-sampling (RUS) as a Data Reduction Method", *Glob. J. Technol. Optim.*, Vol. 01, No. S1, 2016, doi: 10.4172/2229-8711.s1111.

[49] M. Bach, A. Werner, and M. Palt, "The Proposal of Undersampling Method for Learning from Imbalanced Datasets", *Procedia Comput. Sci.*, Vol. 159, pp. 125–134, 2019, doi: https://doi.org/10.1016/j.procs.2019.09.167.

[50] B. Jonathan, P. H. Putra, and Y. Ruldeviyani, "Observation Imbalanced Data Text to Predict Users Selling Products on Female Daily with SMOTE, Tomek, and SMOTE-Tomek", In: *Proc. of 2020 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, pp. 81–85, 2020, doi: 10.1109/IAICT50021.2020.9172033.

[51] P. T. Prasetyaningrum, P. Purwanto, and A. F. Rochim, "Gamification On Mobile Banking

Application: A Literature Review", In: *Proc. of 2022 1st International Conference on Information System & Information Technology (ICISIT)*, pp. 222–227, 2022.

[52] P. L. Skryjomski and B. Krawczyk, "Influence of minority class instance types on SMOTE imbalanced data oversampling", In: *Proc. of Mach. Learn. Res.*, Vol. 74, pp. 7–21, 2017.

[53] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique", *J. Artificial Intell. Res.*, Vol. 16, pp. 321–357, 2002.

[54] R. M. AlZoman and M. J. F. Alenazi, "A Comparative Study of Traffic Classification Techniques for Smart City Networks", *Sensors*, Vol. 21, No. 14, pp. 43–63, 2021, [Online]. Available: http://link.springer.com/10.1007/978-3-319-61313-0

[55] J. Han and M. Kamber, *Data Mining : Concepts and Techniques, Second Edition*, San Francisco: Morgan Kaufmann Publishers, 2006.

[56] S. Wang and X. Yao, "Using class imbalance learning for software defect prediction", *IEEE Trans. Reliab.*, Vol. 62, No. 2, pp. 434–443, 2013, doi: 10.1109/TR.2013.2259203.

[57] M. Khashei, A. Zeinal Hamadani, and M. Bijari, "A novel hybrid classification model of artificial neural networks and multiple linear regression models", *Expert Syst. Appl.*, Vol. 39, No. 3, pp. 2606–2620, 2012, doi: https://doi.org/10.1016/j.eswa.2011.08.116.

[58] S. C. Satapathy, J. V. R. Murthy, P. V. G. D. P. Reddy, B. B. Misra, P. K. Dash, and G. Panda, "Particle swarm optimized multiple regression linear model for data classification", *Appl. Soft Comput.*, Vol. 9, No. 2, pp. 470–476, 2009, doi: https://doi.org/10.1016/j.asoc.2008.05.007.

[59] W. Yao and L. Li, "A new regression model: modal linear regression", *Scand. J. Stat.*, Vol. 41, No. 3, pp. 656–671, 2014.

[60] S. D. Jadhav and H. P. Channe, "Efficient Recommendation System Using Decision Tree Classifier and Collaborative Filtering", *Int. Res. J. Eng. Technol.*, Vol. 3, No. 8, pp. 2113–2118, 2016.

[61] P. Chou, H. H. C. Chuang, Y. C. Chou, and T. P. Liang, "Predictive analytics for customer repurchase: Interdisciplinary integration of buy till you die modeling and machine learning", *Eur. J. Oper. Res.*, Vol. 296, No. 2, pp. 635–651, 2022, doi: 10.1016/j.ejor.2021.04.021.

[62] L. Breiman, "Random Forests", *Int. J. Adv. Comput. Sci. Appl.*, 2001.

[63] L. Wang, *Support vector machines: theory and applications*, Vol. 177. Springer Science & Business Media, 2005.

[64] E. R. Ziegel, *The elements of statistical learning*, Taylor & Francis, 2003.

[65] W. Sun and B. Trevor, "A stacking ensemble learning framework for annual river ice breakup dates", *J. Hydrol.*, Vol. 561, pp. 636–650, 2018, doi: https://doi.org/10.1016/j.jhydrol.2018.04.008.

[66] D. H. Wolpert, "Stacked generalization", *Neural Networks*, Vol. 5, No. 2, pp. 241–259, 1992, doi: 10.1016/S0893-6080(05)80023-1.

[67] E. L. Mencía and F. Janssen, "Learning rules for multi-label classification: a stacking and a separate-and-conquer approach", *Mach. Learn.*, Vol. 105, No. 1, pp. 77–126, 2016, doi: 10.1007/s10994-016-5552-1.

[68] J. Sill, G. Takacs, L. Mackey, and D. Lin, "Feature-Weighted Linear Stacking", *arXiv:0911.0460*, pp. 1–17, 2009, [Online]. Available: http://arxiv.org/abs/0911.0460

[69] T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning", New York, NY, USA: Springer, Vol. 1. 2001.

[70] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System", In: *Proc. of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

[71] C. Y. Ting, N. Y. Z. Tan, H. H. Hashim, C. C. Ho, and A. Shabadin, "Malaysian Road Accident Severity: Variables and Predictive Models", *Lect. Notes Electr. Eng.*, Vol. 603, No. February, pp. 699–708, 2020, doi: 10.1007/978-981-15-0058-9_67.

[72] D. Nielsen, "Tree boosting with xgboost-why does xgboost win "every" machine learning competition?", *NTNU*, 2016.

[73] U. Ependi, A. F. Rochim, and A. Wibowo, "A Hybrid Sampling Approach for Improving the Classification of Imbalanced Data Using ROS and NCL Methods", *Int. J. Intell. Eng. Syst.*, Vol. 16, No. 3, pp. 345–361, 2023, doi: 10.22266/ijies2023.0630.28.