Journal of Applied Science and Emerging Technology (JASET)
Published by Faculty of Applied Science, KMUTNB, Thailand
doi: 10.14416/JASET.KMUTNB.2024.03.002
Vol. 23, No. 3 [2024]: e257236

Research Article

Determining the Distinctiveness of Learners with Outlier Detection Ensembles

Wilawan Inchamnan[1*], Jesada Kajornrit[2], Waraporn Jirapanthong[3]
College of Creative Design & Entertainment Technology, Dhurakij Pundit University Thailand
wilawan.inn@dpu.ac.th [1], jesada.kaj@dpu.ac.th[2], waraporn.jir@dpu.ac.th[3]

ABSTRACT

Gamification is the application of game design principles and elements to non-game contexts. It has been increasingly used to engage and motivate learners. A clear understanding of learners' characteristics is a key success of gamification. Understanding the distinctiveness of the learners' behaviors enables gamification design to encompass all types of participants. This paper then proposes a methodology to assist gamification design in identifying distinctive individual learners within a group. The proposed methodology adopts ensembles of outlier detection techniques to examine how much individual learners differ quantitatively from the group, and utilizes decision tree classifiers to identify the factors contributing to the distinctiveness. The outcomes of these methods are presented in the if-then rules, which assist the interpretability of the discovered insights. This method enables educators and gamification designers to personalize gamified learning environments by focusing on unique learner characteristics.

KEYWORDS: Outliers Detection, Ensembles Method, Decision Trees, Learner Distinctiveness, Gamification Design

*Corresponding Author: Wilawan.inn@dpu.ac.th

Received: 21/06/2024; Revised: 10/10/2024; Accepted: 10/10/2024

1. INTRODUCTION

Gamification is the process of applying game design principles, mechanics, and elements to non-game contexts. The objectives of gamification are to engage users, encourage participation, and motivate desired behaviors (Inchamnan, 2023). Essentially, it involves using elements commonly found in games, such as points, badges, leaderboards, challenges, and rewards, to make activities outside of games more enjoyable and engaging.

Gamification research is a vibrant field with ongoing studies, exploring various aspects of gameful design, motivation, and behavior change. One significant focus is on integrating gamification elements into education in order to enhance learner engagement and opportunities within educational settings. This approach involves understanding how elements like points, badges, and leaderboards can influence behavior and mindset. Objective is aim to improve instructional activities and foster better learning outcomes (Inchamnan, & Chomsuan, 2021).

The history of integrating game elements into daily life has evolved into modern gamification practices. The studies explore its evolution, motivational effects, productivity enhancements, and impacts on learning methodologies (Sharma & Sharma, 2023). Furthermore, gamification research aims to bridge the gap between gamification design and human motivation. It also offers valuable insights for researchers and practitioners in the field of information technology design (Tang & Zhang, 2019).

Presently, data analytics has become an important issue in several areas, and gamification is no exception. Combining gamification with data analytics allows organizations not only to create engaging experiences but also to gather valuable data on user interactions, preferences, and performance. For example, a company might use gamification techniques in its employee training program and then use data analytics to assess which modules are most effective based on employee engagement and learning outcomes (Greene, 2023). On the other hand, data analytics enables gamification designers to understand their learners in advance and design personalized content (Stone et. al., 2015).

The integration of gamification with data analytics can be accomplished for several purposes (Poolsawas & Niranatlamphong, 2017), for examples, behavior tracking, performance evaluation, personalization, iterative improvement and predictive analytics. Personalization is interesting point for our study. It could be satisfied to all learners if the contents in the gamification math to everybody. To accomplish this mission, gamification designers have to understand the distinctiveness of the learners.

If we define the distinctiveness as the measure of learner's uniqueness, it could be hypothesized that most learners exhibit common behaviors, while some learner deviate from the norm. In the area of data analytics, there is one type of analysis named outlier detection that tries to identify data points or observations that deviate significantly from the rest of the dataset (Chandola et. al., 2009). These outliers may indicate anomalies or unique patterns within the data. Imagine that if some deviated learners in the class are similar to the unique patterns, it would be very challenging to design gamification content to serve those learners. However, we have to find those unique learners first. The purpose of this paper is to present an interpretable approach to apply outlier detection technique to discover the distinctiveness of learners.

The rest of the paper is organized as follows. Section 2 provides some outlier detection methods used in this study. The information of gamification data is briefly discussed in section 3. Section 4 proposes our methodology, while the experimental results and analysis are discussed in the section 5. Finally, section 6 is the conclusion.

2. SELECTED OUTLIER DETECTION TECHNIQUES

Outlier detection is a data analysis technique aimed at identifying observations or data points that deviate significantly from the majority of the data. These deviations, known as outliers, can signal anomalies, errors, or rare occurrences within the dataset. Over recent decades, a variety of techniques have been developed for outlier detection, including classification methods, clustering methods, k-nearest neighbor methods, and statistical methods. Additionally, these techniques can be categorized into supervised, semi-supervised, and unsupervised methods.

In practice, one of the recommended libraries for outlier detection is PyOD (Han et. al., 2022). PyOD is a versatile Python library for detecting outliers in multivariate data. It offers more than fifty algorithms that have been successfully used in numerous academic research studies and commercial products. In this study, the authors initially adopted thirteen unsupervised algorithms in the proposed method. However, this does not imply that the proposed method is limited to these algorithms for outlier detection. The adopted algorithms are summarized as follows.

- Principal Component Analysis (PCA) (Shyu et. al., 2003) is a linear dimensionality reduction using singular value decomposition of the data to project it to a lower dimensional space. When used for outlier detection, it projects the data to the lower dimensional space and then uses the reconstruction errors as the anomaly scores.
- One-Class Support Vector Machine (OCSVM) (Schölkopf et. al., 1999) maximizes the margin between the origin and the normal samples, and defines the decision boundary as the hyperplane that determines the margin.
- Local Outlier Factor (LOF) (Breunig et. al., 2000) measures the local deviation of the density of a given sample with respect to its neighbors. It is local in that the anomaly score depends on how isolated the object is with respect to the surrounding neighborhood.
- Clustering Based Local Outlier Factor (CBLOF) (He et. al., 2003) calculates the anomaly score by first assigning samples to clusters, and then using the distance among clusters as anomaly scores.
- Connectivity-Based Outlier Factor (COF) (Tang et. al., 2002) uses the ratio of the average chaining distance of data points and the average chaining

distance of k-th nearest neighbor of the data point, as the anomaly score for observations.

- Histogram- Based Outlier Detection (HBOS) (Goldstein & Dengel, 2012) assumes feature independence and calculates the degree of outlyingness by building histograms.
- K-Nearest Neighbors (KNN) (Ramaswamy et. al., 2000) views the anomaly score of the input instance as the distance to its k-th nearest neighbor. Subspace Outlier Detection (SOD) (Kriegel et. al., 2009) aims to detect outliers in varying subspaces of high-dimensional feature space.
- Copula Based Outlier Detector (COPOD) (Li et. al., 2020) is a hyperparameter-free, highly interpretable outlier detection algorithm based on empirical copula models.
- Empirical-Cumulative-distribution-based Outlier Detection (ECOD) (Li et. al., 2022) is a hyperparameter-free, highly interpretable outlier detection algorithm based on empirical cumulative distribution function (ECDF). Basically, it uses ECDF to estimate the density of each feature independently, and assumes that outliers locate the tails of the distribution.
- Gaussian Mixture Model (GMM) (Zong et. al., 2018) is a probabilistic model used for clustering and density estimation tasks. It is a parametric model that represents the probability distribution of a dataset as a weighted sum of several Gaussian distributions.
- Lightweight on-line detector of anomalies (LODA) (Pevný, 2016) is an ensemble method and is particularly useful in domains where a large number of samples need to be processed in real-time or in domains where the data stream is subject to concept drift and the detector needs to be updated online.
- Isolation Forest (IForest) (Liu et. al., 2008) observations by randomly selecting a feature and then randomly selecting a split value between the maximum and minimum values of the selected feature.

As this paper does not focus extensively on outlier detection algorithms, a brief overview of selected algorithms is just provided. For readers interested in a deeper exploration of this topic, the authors recommend consulting references (Chandola et. al., 2009), (Aggarwal, 2017), and (Mehrotra et. al., 2017) for more details.

Outlier detection algorithms mentioned above have their own advantages. The different techniques could lead to differences in detection outcomes. Thus, the goal is to combine these advantages to increase detection confidence. Ensemble techniques are not a new concept; they have been successfully used in various application domains (Mehrotra et. al., 2014). In the current study, we use an ensemble technique to detect distinctive behaviors of learners in the gamification domain. Such a problem is essentially equivalent to outlier detection.

As our dataset do not have label, the proposed methodology then falls in the unsupervised analysis. It is difficult to apply supervised classification algorithms directly to find the outliers. However, using ensemble of unsupervised algorithms can enhance prediction confidence by focusing on students whose behaviors qualitatively deviate from the group norm, signaling potential risks or opportunities for intervention (Novoseltseva, 2022).

3. GAMIFICATION DATASET

This study gathered data from a survey completed by 93 students in the class. The dataset includes information about the students (No.1 - 4) and their answers to 20 questions (No. 5 - 24). Table 1 provides a summary of the dataset details. Table 2 exhibits the possible values of category data. Table 3 shows statistics of the dataset.

Table 1 Question in gamification dataset

No	Feature	Data type	Name
1	Sex	Category (2 values)	sex
2	Age	Numeric [17, 33]	age
3	GPA	Numeric	gpa

		[0.00, 4.00]	
4	Faculty	Category (5 values)	faculty
5	What type of game do the students like to play the most?	Category (9 values)	game
6	Can students reflect their experiences, interests, and familiar perspectives while studying in the classroom?	Ordinal [1, 5] 1 = Poor , 5 = Excellent	q1
7	Do students observe or participate in unfamiliar things? Are they exposed to new situations or content they haven't encountered before?	Ordinal [1, 5] 1 = Poor , 5 = Excellent	q2
8	Can students identify, think, discuss, and collaborate with others on the content of the lesson?	Ordinal [1, 5] 1 = Poor , 5 = Excellent	q3
9	Do students understand the content according to the step-by-step theory taught?	Ordinal [1, 5] 1 = Poor , 5 = Excellent	q4
10	Can students analyze case studies assigned by the instructor?	Ordinal [1, 5] 1 = Poor , 5 = Excellent	q5
11	Do students understand their own and others' perspectives, interests, and motivations?	Ordinal [1, 5] 1 = Poor , 5 = Excellent	q6
12	Can students apply newly acquired knowledge in real-life situations at present?	Ordinal [1, 5] 1 = Poor , 5 = Excellent	q7
13	Can students design innovations, create new things, or transmit their own learning?	Ordinal [1, 5] 1 = Poor , 5 = Excellent	q8
14	Do students understand the rules of teaching and learning?	Ordinal [1, 5] 1 = Poor , 5 = Excellent	q9
15	Do students understand the objectives of each learning session?	Ordinal [1, 5] 1 = Poor , 5 = Excellent	q10
16	Each time scoring in the room, De students feel like coming to study and feel rewarded?	Ordinal [1, 5] 1 = Poor , 5 = Excellent	q11
17	Do students prefer teachers to provide current challenging case studies?	Ordinal [1, 5] 1 = Poor , 5 = Excellent	q12
18	Do students enjoy competing both inside and outside the classroom and strive to be the best in the classroom?	Ordinal [1, 5] 1 = Poor , 5 = Excellent	q13
19	Are the classroom activities fun and appropriate?	Ordinal [1, 5] 1 = Poor , 5 = Excellent	q14
20	Do the students prefer learning online more than in the classroom?	Ordinal [1, 5] 1 = Poor , 5 = Excellent	q15
21	Do instructors prefer setting deadlines for assignments, and if exceeded, do they deduct points?	Ordinal [1, 5] 1 = Poor , 5 = Excellent	q16
22	Do students like working in groups or not?	Ordinal [1, 5] 1 = Poor , 5 = Excellent	q17
23	What type of game character the students are the most?	Category (4 values)	q18
24	What personal type awareness of students?	Category (5 values)	q19

Table 2 Possible values of the category data

No.	Possible Values
1	(1) Male (2) (2) Female
4	(1) Information Technology (2) Tourism and Hospitality (3) Business Management (4) Integrative Medicine (5) Communication Arts
5	(1) Role-playing (2) Shooting (3) Fighting (4) Adventure

	(5) Strategy (6) Simulation (7) Puzzle (8) Sport (9) Not playing
23	(1) Explorer: I enjoy learning to gain knowledge in new things continuously. (2) Killer: I like to be number 1 or at least at the beginner level in the classroom. (3) Achiever: I like to follow the teacher's instructions thoroughly according to the guidelines provided by the teacher. (4) Socializer: I like to be a part of my friends in the classroom because it feels fun.
24	(1) I don't like competition and be afraid of making mistakes. (2) I focus on practical implementation to see a clear picture. (3) I am a confident person who emphasizes interaction. (4) I prefer working alone and don't enjoy socializing. (5) Other

Table 3 Statistics of the Dataset

No	Statistics
1	(1) = 27, (2) = 66
2	Range = [17, 33], Average = 21.5
3	Range = [0.00, 4.00], Average = 3.36
4	(1) = 14, (2) = 1, (3) = 12, (4) = 65, (5) = 1
5	(1) = 7, (2) = 24, (3) = 8, (4) = 6, (5) = 4, (6) = 7, (7) = 7, (8) = 4, (9) = 26
6	(1) = 0, (2) = 4, (3) = 24, (4) = 40, (5) = 25
7	(1) = 0, (2) = 3, (3) = 28, (4) = 41, (5) = 21
8	(1) = 0, (2) = 2, (3) = 16, (4) = 42, (5) = 33
9	(1) = 0, (2) = 1, (3) = 28, (4) = 35, (5) = 29
10	(1) = 0, (2) = 2, (3) = 26, (4) = 43, (5) = 22
11	(1) = 0, (2) = 1, (3) = 30, (4) = 36, (5) = 26
12	(1) = 1, (2) = 0, (3) = 29, (4) = 38, (5) = 25
13	(1) = 1, (2) = 2, (3) = 38, (4) = 26, (5) = 26
14	(1) = 0, (2) = 0, (3) = 16, (4) = 31, (5) = 46
15	(1) = 1, (2) = 1, (3) = 26, (4) = 29, (5) = 36
16	(1) = 5, (2) = 7, (3) = 29, (4) = 32, (5) = 20
17	(1) = 2, (2) = 5, (3) = 29, (4) = 34, (5) = 23
18	(1) = 20, (2) = 14, (3) = 36, (4) = 10, (5) = 13
19	(1) = 0, (2) = 10, (3) = 27, (4) = 32, (5) = 24
20	(1) = 18, (2) = 14, (3) = 33, (4) = 10, (5) = 18
21	(1) = 6, (2) = 5, (3) = 32, (4) = 23, (5) = 27
22	(1) = 19, (2) = 5, (3) = 40, (4) = 15, (5) = 14
23	(1) = 60, (2) = 4, (3) = 14, (4) = 15
24	(1) = 26, (2) = 34, (3) = 17, (4) = 5, (5) = 13

4. PROPOSED METHODOLOGY

The proposed methodology is shown in the Figure 1 The methodology consists of five steps, details as follows:

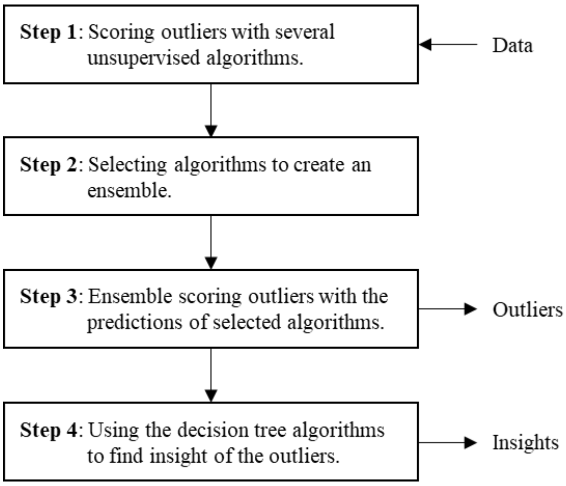


Figure 1 The proposed methodology

Step 1: Scoring outliers with various selected algorithms. In this step, many algorithms are applied to the data in order to assign outlier scores to the data points. It is, of course, evident that the choice of algorithms results in differing scoring outcomes. The scoring outcome of each algorithm is normalized within a range between 0 and 1. The methodology does not limit the number of algorithms. In our experiment, however, we utilize thirteen algorithms provided in the PyOD library (Han et. al., 2022) mentioned in the section 2 of this paper. The source of the algorithms can be found in (Zhao, 2024).

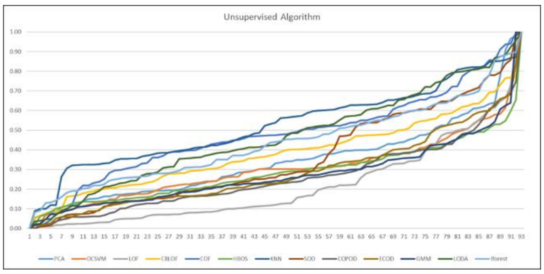
Step 2: Selecting algorithms to create an ensemble. In this step, a subset of algorithms is chosen from all available options. The outlier scores generated by each algorithm in step 1 are sorted from 0 to 1. The sorted outliner scores (y-axis) are plotted against the number of data points (x-axis). To select the algorithms, we observe at 90% of the total number of data points. At this juncture, the chosen algorithms should encompass all outlier scores ranging from the minimum to the maximum scores.

Step 3: Scoring outliers with the predictions of algorithms. It is assumed that data points are identified as outliers if they fall within the top 10% of data points with the highest outlier scores. Through various algorithms, certain data points may be labeled as normal (label = 0) by some algorithms and as outliers (label = 1) by others. Therefore, the ensemble score is defined as the sum of all selected algorithms (label). For example, if three out of five algorithms predict a data point as an outlier, then the ensemble score is 3.

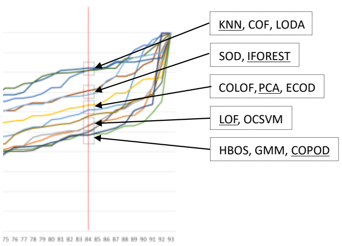
Step 4: Utilizing the decision tree algorithm to investigate insights. Once the ensemble score has been assigned to the data points, it can be used as the label for supervised learning. The ensemble label reflects the distinctive level of the data points. The higher the distinctive level, the further the distance from normal behaviors. Supervised machine learning techniques such as decision trees can be employed to identify the causes of distinctiveness. To clearly illustrate the proposed methodology, the gamification data will be processed through all steps mentioned above.

5. EXPERIMENTAL RESULTS

According to the step 1, the data are preprocessed into the machine learning-suitable format. Numeric features "age" and "gpa" as well as ordinal features "q1" to "q17," are normalized using min-max scaling to a range of 0 to 1. Categorical features "sex", "game", "faculty", "q18", and "q19" are encoded using the one-hot method. Therefore, the final number of features is 45. Next, the preprocessed dataset is applied with thirteen unsupervised outlier detection algorithms, that is, PCA, OCSVM, LOF, CBLOF, COF, HBOS, KNN, SOD, COPOD, ECOD, GMM, LODA, and, IForest.



(a)



(b)

Figure 2 (a) The plot of thirteen normalized outlier scores, and (b) Selected algorithms at most 10 percent datapoints

According to step 2, the outlier scores of each algorithm are sorted from minimum to maximum. Then they are normalized to a range of 0 to 1 for comparison. Figure 2 shows the plot of those outlier scores. The x-axis represents the data points, and the y-axis represents the normalized scores. One can observe that due to differences in algorithms, variations in scoring occur. Some data points are detected as outliers by one algorithm but as normal by other algorithms.

To select the algorithms, we observe at 90% of the total number of data points, in this case, at 84 on the x-axis. At this juncture, we select five algorithms that encompass the entire outlier scores range (see Figure 2). The selected algorithms are KNN, IForest, PCA, LOF, and COPOD. The selection is rather arbitrary, thus other ones may select difference set of algorithms. According to step 3, the data points belonging to the top 10 percent of the highest outlier scores are predicted as outliers (class 1 or score = 1); otherwise, they are predicted as normal (class 0 or score = 0). Therefore, if we sum the scores from the selected algorithms, each data point may have a score ranging from 0 to 5. This is referred to as the ensemble score. A score of 0 means that no algorithm predicts the data point as an outlier (extremely normal), whereas a score of 5 means that every algorithm predicts the data point as an outlier (extremely outlier). In our experiment, the number of data points are in Table 4.

Table 4 The numbers of data in ensemble scores.

Score	Number of data point	Percent of data point
0	67	72.04%
1	12	12.90%
2	7	7.52%
3	4	4.30%
4	2	2.15%
5	1	1.07%
sum	93	100%

We can use this score as the distinguishing level for students. 67 students in level 0 are considered normal students. This largest group of students can be utilized to design gamification for general purposes. The next groups consist of 12 and 7 students in levels 1 and 2, respectively, who differ from the normal students and can be grouped together for specific purposes. Finally, 4 and 2 students in levels 3 and 4 are distinctive students who may serve very specific purposes in gamification design. 1 student in level 5 may represent noise in the data.

According to step 4, We label the data with score derived in the step 3 and utilize the Decision Tree classification algorithm to generate decision tree diagram for analysis. The experiment generates decision tree with three criteria including gain ration (Figure 3), information gain (Figure 4), and gini index (Figure 5) to cooperative analysis. The experiment set parameter depth of tree equal to 5 for interpretable and visualizable purposes.

According to the Figure 3, it could be observed from the tree that:

- IF GPA less than 0.5 THEN the student data probably is noise data.
- IF GPA greater than 0.5 AND student tend to focus on practical implementation to see a clear picture THEN student tend to have high distinctiveness from normal student. And if that student prefers a fighting game, he/she have relative higher distinctiveness.

According to the Figure 4, it could be observed from the tree that:

- IF student tend to focus on practical implementation to see a clear picture AND student like fighting game THEN student tend to have high distinctiveness from normal
- IF student is not a good in apply newly acquired knowledge in real-life situations at present AND the student prefers shooting game THEN the student tend to have high distinctiveness from others.

According to the Figure 5, it could be observed from the tree that:

- IF student tend to focus on practical implementation to see a clear picture AND student like fighting game THEN student tend to have high distinctiveness from normal

From the three models, a common pattern that leads to high distinctiveness is in the second selection "student who tends to focus on practical implementation to see a clear picture" AND "student who prefers fighting game" of the questions named "q19" and "game", respectively. This insight enables gamification designer to have more information to analyze high distinctives students from this analysis.

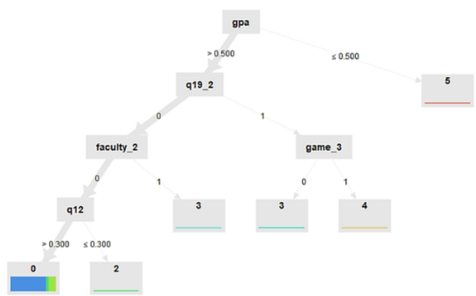


Figure 3. The decision tree modeled with gain ratio criterion

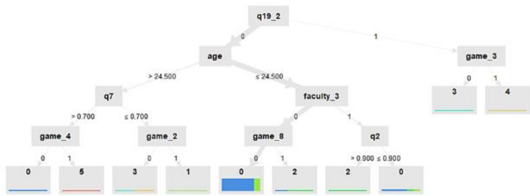


Figure 4. The decision tree model with information gain criterion

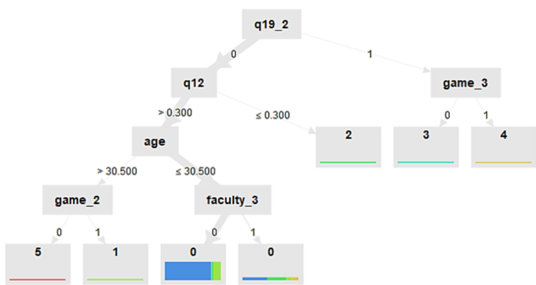


Figure 5 The decision tree modeled with gini index criterion

Understanding learners' behaviors can significantly enhance gamification design. The proposed method offers an alternative approach for discerning learners' distinctiveness through If-Then rules. However, this paper represents preliminary research and provides only a conceptual framework for key takeaways. Future research should involve a larger sample size to strengthen the study's validity. Additionally, as game technology evolves, new game types, such as soul-like games, have been defined, along with new gamer classifications, exemplified by the 9 Quantic gamer types (Greene, 2024). Furthermore, the exploration of new data analytics methods for gamification design remains an ongoing area of investigation.

6. CONCLUSION

Understanding the characteristics of learners makes gamification design more attractive and effective. Data analytics enables gamification designers to better understand learners. This study proposes an approach to data analytics named outlier detection to discover the distinctiveness of learners. The study defines distinctiveness as the learners' characteristics that are far from the norm of the group. The proposed method consists of using an ensemble of outlier detection algorithms to score the distinctiveness of the learners and using decision tree algorithms to extract decision rules of distinctiveness. The study exemplifies the proposed method with data collected from one class and shows how to discover insights from the data. However, machine learning has limitations in addressing imbalanced datasets. Some research proposes a framework that combines outlier detection and feature reduction to improve model performance, highlighting the importance of outliers in identifying rare or distinctive instances, which ultimately enhances classification accuracy (Lusito et. al., 2024). The outcomes of the methodology are expressed in the form of IF-THEN rules, which are easily interpretable by humans. Gamification analysts can utilize the extracted insights to tailor their instructional strategies more effectively to specific learner groups. In our study, students who exhibit a preference for fighting games and focus on practical implementation to gain a clearer understanding were identified as distinctive learners. This observation raises questions for gamification analysts, particularly regarding the underlying reasons for these preferences, which may warrant further experimentation. Future research should aim to gather additional data to reinforce our findings. The integration of gamification design with outlier detection remains an intriguing area of study.

Outlier detection doesn't inherently judge whether the distinctiveness is positive or negative. It simply identifies students whose behaviors, preferences, or performances differ from the majority. The interpretation of this distinctiveness, for example, good or bad learners, depends on the context in which these outliers are analyzed. For example, a student with a unique learning style that differs from the typical approach may require personalized learning strategies that could enhance their performance, which would be beneficial once addressed (Novoseltseva, 2022).

REFERENCES

Aggarwal, C. C. (2017). *Outlier analysis*, 2nd ed., Bangkok, Thailand: Fophrinting Chamkat. pp.113-140.

Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000). LOF: identifying density-based local outliers. *In Proceedings of the 2000 ACM SIGMOD international conference on Management of data* (pp. 93-104).

Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey, *ACM computing surveys*, *41*(3).

Goldstein, M., & Dengel, A. (2012). Histogram-based outlier score (hbos): A fast unsupervised anomaly detection algorithm. KI-2012: *Poster and Demo Track of the 35th German Conference on Artificial Intelligence (KI-2012)* (pp. 59-63).

Greene, K. (2023). Big Data and Gamification: How to Play It Smart. Retrieved January, 2023, from https://emfluence.com/blog/big-data-and-gamification-how-to-play-it-smart

Greene, K. (2024). The 9 Quantic Gamer Types, 2024, from https://quanticfoundry.com/gamer-types/

Han, S., Hu, X., Huang, H., Jiang, M., & Zhao, Y. (2022). Adbench: Anomaly detection benchmark. Advances in Neural Information Processing Systems, *36th Conference on Neural Information Processing Systems (NeurIPS) Track on Datasets and Benchmarks* (pp. 32142-32159).

He, Z., Xu, X., & Deng, S. (2003). Discovering cluster-based local outliers. *Pattern recognition letters*, *24*(9-10), 1641-1650.

Inchamnan, W. (2023). Gamification Pedagogy Trigger Design for Creative Experiences in Metaverse Class Room. *Journal of Applied Science and Emerging Technology (JASET)*, *22*(2), 251539.

Inchamnan, W., & Chomsuan, J. (2021). The Gamification Design for Affordances Pedagogy. *Advances in Science, Technology and Engineering Systems Journal*, *6*(4), 138-146.

Kriegel, H. P., Kröger, P., Schubert, E., & Zimek, A. (2009). Outlier detection in axis-parallel subspaces of high dimensional data. *In Advances in Knowledge Discovery and Data Mining: 13th Pacific-Asia Conference, PAKDD 2009 Bangkok, Thailand, April 27-30, 2009 Proceedings 13* (pp. 831-838). Springer Berlin Heidelberg.

Li, Z., Zhao, Y., Botta, N., Ionescu, C., & Hu, X. (2020). COPOD: copula-based outlier detection. *In 2020 IEEE international conference on data mining (ICDM)* (pp. 1118-1123). IEEE.

Li, Z., Zhao, Y., Hu, X., Botta, N., Ionescu, C., & Chen, G. H. (2022). Ecod: Unsupervised outlier detection using empirical cumulative distribution functions. *IEEE Transactions on Knowledge and Data Engineering*, *35*(12), 12181-12193.

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. *In 2008 eighth ieee international conference on data mining* (pp. 413-422). IEEE.

Lusito, S., Pugnana, A., & Guidotti, R. (2024). Solving imbalanced learning with outlier detection and features reduction. *Machine Learning*, *113*(8), 5273-5330. https://doi.org/10.1007/s10994-023-06448-0

Mehrotra, K. G., Mohan, C. K., Huang, H., Mehrotra, K. G., Mohan, C. K., & Huang, H. (2017). *Anomaly detection*, Springer International Publishing. pp. 21-32.

Novoseltseva, D. (2022). Outlier Detection as a Tool for Reinforcing Data Analysis and Prediction in Education [Doctoral's thesis]. Université Paul Sabatier-Toulouse III.

Pevný, T. (2016). Loda: Lightweight on-line detector of anomalies. *Machine Learning*, *102*, 275-304.

Poolsawas, B., & Niranatlamphong, W. (2017). Using a Game Development Platform to Improve Advanced Programming Skills. *Journal of Reviews on Global Economics*, *6*, 328-334.

Ramaswamy, S., Rastogi, R., & Shim, K. (2000). Efficient algorithms for mining outliers from large data sets. *In Proceedings of the 2000 ACM SIGMOD international conference on Management of* data (pp. 427-438).

Schölkopf, B., Williamson, R. C., Smola, A., Shawe-Taylor, J., & Platt, J. (1999). Support vector method for novelty detection. *Advances in neural information processing systems*, *12*.

Sharma, D., & Sharma, J. (2023, July). Evolution of gamification, its implications, and its statistical impact on the society. *In International Conference on Emerging Trends in Design & Arts* (pp. 8-20).

Shyu, M. L., Chen, S. C., Sarinnapakorn, K., & Chang, L. (2003, November). A novel anomaly detection scheme based on principal component classifier. *In Proceedings of the IEEE foundations and new directions of data mining workshop* (pp. 172-179). Piscataway, NJ, USA: IEEE Press..

Stone, J., Sanyal, J., Castello, C., & New, J. (2015). Gamification and Visualization of Sensor Data Analysis in Research Buildings. *Oak Ridge National Laboratory (ORNL)*.

Tang, J., & Zhang, P. (2019). Exploring the relationships between gamification and motivational needs in technology design. *International Journal of Crowd Science*, *3*(1), 87-103.

Tang, J., Chen, Z., Fu, A. W. C., & Cheung, D. W. (2002). Enhancing effectiveness of outlier detections for low density patterns. *In Advances in Knowledge Discovery and Data Mining: 6th Pacific-Asia Conference, PAKDD 2002 Taipei, Taiwan, May 6-8, 2002 Proceedings 6* (pp. 535-548). Springer Berlin Heidelberg.

Zhao, Y. (2024). Welcome to PyOD documentation!. Retrieved January, 2024, from https://pyod.readthedocs.io/en/latest/

Zimek, A., Campello, R. J., & Sander, J. (2014). Ensembles for unsupervised outlier detection: challenges and research questions a position paper. *Acm Sigkdd Explorations Newsletter*, *15*(1), 11-22.

Zong, B., Song, Q., Min, M. R., Cheng, W., Lumezanu, C., Cho, D., & Chen, H. (2018). Deep autoencoding gaussian mixture model for unsupervised anomaly detection. *In International conference on learning representations*.