Optimizing Random Forest Algorithm to Classify Player's Memorisation via In-game Data

Akmal Vrisna Alzuhdi [a,1], Harits Ar Rosyid [a,2,*], Mohammad Yasser Chuttur [b,3], Shah Nazir [c,4]

[a] Department of Electrical Engineering and Informatics, Universitas Negeri Malang
Jl. Semarang no. 5, Malang 65145, Indonesia
[b] Department of Software and Information Systems, University of Mauritius
2nd floor Phase II building University of Mauritius Reduit MU, 80837, Mauritius
[c] Department of Computer Science, University of Swabi
Ambar, Swabi, Khyber Pakhtunkhwa 94640, Pakistan
[1] akmal.vrisna.1705356@students.um.ac.id; [2] harits.ar.ft@um.ac.id; [3] y.chuttur@uom.ac.mu;
[4] nasirshah@uoswabi.edu.pk
* corresponding author

ARTICLE INFO

ABSTRACT

Assessment of a player's knowledge in game education has been around for some time. Traditional evaluation in and around a gaming session may disrupt the players' immersion. This research uses an optimized Random Forest to construct a non-invasive prediction of a game education player's Memorization via in-game data. Firstly, we obtained the dataset from a 3-month survey to record in-game data of 50 players who play 4-15 game stages of the Chem Fight (a test case game). Next, we generated three variants of datasets via the preprocessing stages: resampling method (SMOTE), normalization (min-max), and a combination of resampling and normalization. Then, we trained and optimized three Random Forest (RF) classifiers to predict the player's Memorization. We chose RF because it can generalize well given the high-dimensional dataset. We used RF as the classifier, subject to optimization using its hyperparameter: n_estimators. We implemented a Grid Search Cross Validation (GSCV) method to identify the best value of n_estimators. We utilized the statistics of GSCV results to reduce the weight of n_estimators by observing the region of interest shown by the graphs of performances of the classifiers. Overall, the classifiers fitted using the BEST n_estimators (i.e., 89, 31, 89, and 196 trees) from GSCV performed well with around 80% accuracy. Moreover, we successfully identified the smaller number of n_estimators (OPTIMAL), at least halved the BEST n_estimators. All classifiers were retrained using the OPTIMAL n_estimators (37, 12, 37, and 41 trees). We found out that the performances of the classifiers were relatively steady at ~80%. This means that we successfully optimized the Random Forest in predicting a player's Memorization when playing the Chem Fight game. An automated technique presented in this paper can monitor student interactions and evaluate their abilities based on in-game data. As such, it can offer objective data about the skills used.

I. Introduction

A game is a work of art based on specific rules. These rules drive the end of the game based on the player's actions within the game. A player should use the tools and objects provided in the game to achieve victory. Entertainment is paramount in games, but it is a potential vehicle for training and education through basic thinking skills to solve conflicts or problems [1][2]. Educational games integrate complex principles, including knowledge, pedagogy, decision-making, collaboration, and gaming [3]. The primary aim of an educational game is learning and having fun in unity [4].

Presenting knowledge in the educational game is often wrapped or decorated at the game level. For instance, the Number Munchers game (popular in the 1980s and 1990s) [5] represents dots with math equations. The player aims to collect equations that produce a particular answer (the mission objective). Here, players can learn math equations full of joy (e.g., an experience after evading an

enemy). Number Muncher game showcases an educational game with learning and gaming elements that make it fun and motivating [6]. In addition, by playing an educational game, players can repetitively play a game level (learning tasks as the mission objective) he failed. This repetitiveness in educational games hinders the fear of losing marks.

Using bibliometric analysis has allowed for the investigation of serious gaming research trends [7][8][9][10][11]. The data showed a rise in serious game publications in recent years, highlighting the growing significance of serious games in education. Many academic fields, including education technology, psychology, the medical sciences, the environmental sciences, and corporate economics, have studied serious games. Research has also focused on the use of serious games to help persons with disabilities [9], with an emphasis on education and computer games as the most popular game genres and game platforms, respectively. Collecting data for serious game analytics has proven difficult, with pre-game, in-game, and post-game data being the most common. Digital games and gamification have proven helpful in nursing education in fostering active involvement, elevating satisfaction levels, and imparting skills [10].

There are complex experiences of the player involved during game playing. Affective experiences were reported to exist in an educational game, such as Emotion [12], Motivation [13][14][15][16], and Enjoyment [17][18]. Those articles prove that affective experiences can aid the same significance as the learning goal. However, the knowledge of the player reportedly dominated this research topic, such as Travel in Europe and Sea Game [17][19], Math games [20], Crystal Island Narrative-based game [21][22], and many more. Serious game players are more empowered to become more actively involved, not only in the learning process but also in the design and development of cutting-edge formative assessment tools, as discussed by Hainey et al., and serious games are becoming more and more popular as alternative supplementary learning approaches across all disciplines at all levels [23].

Hence, assessing a player's knowledge due to playing an educational game is the core. An assessment based on one, two, or only a small number of the game's attributes or indicators (e.g., final result, total failures, or duration) sometimes needs to be clarified as to what so-called learning is. For instance, a victory in a game session may indicate whether the player has understood the knowledge with commitment, through a lucky guess, or just playing around. To solve such a problem, one can apply a traditional assessment method via a questionnaire or pre- and post-game exams. This is undoubtedly a reliable and effective assessment method for a user's learning [24]. The numeric difference between pre and post-game examinations quantitatively measures the learning gain. Traditional assessment via a test within a game session may disrupt the enjoyment. For instance, a questionnaire to self-report an affective experience, such as enjoyment [25]. However, these assessment methods are interrupting the gaming experience. The players must dismiss the exciting gaming experience to a seriously thoughtful test. Not many players can deal with such conditions that can lead to disengagement from the game. Behavior observation of the players is not practical as well because of the subjectiveness of the observer.

Meanwhile, there exist in-game actions and the corresponding game level (as inputs) that can represent an experience (as an output) [26][27][28]. However, how to identify and what information is relevant to the player's learning is proprietary to the game. More importantly, optimally correlating the input values and output is the goal of this research. Once we optimally train the prediction model, it should non-intrusively and accurately assess the player's knowledge while maintaining immersion in gaming [28]. Considering the vast amount of information one can retrieve from a game; a data mining approach should fit the task to solve the problem at hand [29][30][31][32]. Say the player's knowledge is categorized as memorization type. A classification technique can solve that, such as [33]. A potential solution is implementing a Random Forest classification to predict the player's Memorization since it is robust with high dimensional, high accuracy, and good generalization [34]. The data often acquired from human players is unreliable and generally imbalanced. Optimally categorizing the player's Memorization should provide an unbiased evaluation that is important for customized learning experiences, adaptable game mechanics, or customized feedback to improve learning outcomes. Thus, we need an optimization method for the Random Forest classifier to handle such data. To accommodate that, we want to

experiment with datasets that preprocess the dataset using a resampling method via SMOTE, normalizing the data using the min-max method, and combining both preprocessing. We expect that our approach is replicable in other educational games since the procedures are straightforward and clear. With the seamless prediction of the player's Memorization, we can identify more insights about the correlation between gaming actions and learning experience. The more variety of game education with seamless assessment can lead to standard in-game data that contribute to learning experiences. Thus, it provides a reliable guideline for designing educational games based on the most relevant in-game data.

The following section discusses the proposed methodology to develop an in-game assessment of the player's knowledge when playing an educational game. It starts by describing the test case educational game and follows the proposed methodology, including the data collection method and optimization experiment.

II. Method

A. Overview of the educational game as the case study

This study uses a game called Chem Dungeon (Chemical Dungeon) as the test case [35]. It is an educational game in introductory chemistry that helps players memorize atoms and chemical compounds. The game genre of Chem Dungeon is a roguelike in a labyrinth. Chem Dungeon's labyrinth comprises paths, walls, intersections, and dead-end alleys (Figure 1) reproduced from [35]. An avatar starts from a spawn point and then collects and forms a chemical compound to reveal an escape portal at the bottom-right of the labyrinth. The avatar should evade Non-Player Characters (NPCs) and avoid constructing incorrect chemical compounds. The avatar has an atomic shield (an atom ready for bonding with others), and details of the atom are readable near the spawn point. When the avatar strikes an atomic mine (blue shield), the game informs the compound-forming result or atom properties readable at the top center of the labyrinth. Game attributes on the right side of the maze include lives (heart icon), experience in a red bar, the remaining ammunition (number), total bonds made (number), and the countdown timer. Inside the labyrinth, there are bullets (yellow item), atoms (blue item), and live potions (red item) that are collectible for the avatar. Each collection of bullets increases the ammunition for the avatar. A live potion can restore the avatar's life.
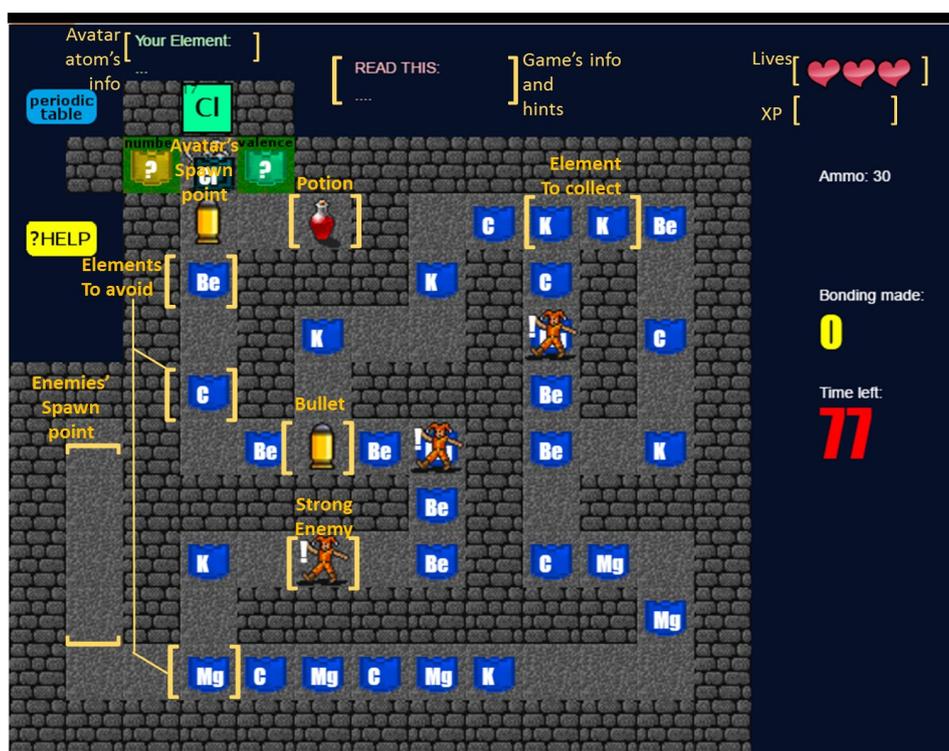

Fig 1. Chem dungeon game

Chem Dungeon's goals are to find the right element to create a compound and pass through the escape portal within 90 seconds. The avatar initially spawns in its residence, and the NPCs start in the diagonal pathways of the labyrinth (bottom-left to top-right). Players can press keyboard keys *a, s, d,* and *w* to navigate the avatar to the left, right, up, and down. The character must stay clear of NPCs and atomic mines while exploring the labyrinth. It loses one life whenever it collides with a weak opponent or a bad atomic mine. The character can also shoot an atomic mine to clear the way. When the bullet strikes a strong opponent, it changes that enemy's state to one of weakness (white-colored NPC). Then, the avatar can capture a weak NPC to make it respawn at its house, opening a new path for the avatar. As a result, the avatar can search for and gather the appropriate element (mine), creating a compound with other atoms. Hence, a piece of educational information on the chemical compound appears at this time. As a result, this game condition should engage participants to memorize and understand the learning materials. The escape portal opens once the avatar has gathered the correct atom ten times. Finally, the avatar receives a Victory by passing through the escape portal—otherwise, a defeat results from losing all lives or running out of time.

The Chem Dungeon game contains 100 chemical compounds constructed by at least two atoms. A compound is shown as character strings representing the symbol, name, and bonding atoms. For instance, two Hydrogen and one Oxygen atoms construct an H_2O representing the water compound. In the game, an atom is a collectible object as an atomic symbol, e.g., C, Ag, N, if more than one atom of the same type appears as a combination of the total atom and the atomic symbol, e.g., 2H, 6B.

The information that follows provides some useful game-playing advice. Although each game has different element options, the objective is to create a single compound (repeatedly). Players new to the game frequently use a trial-and-error approach and are completely conscious of not wasting their remaining lives. The player should, therefore, attentively peruse the text communication corresponding to the compound-forming effort's most recent outcome. Every time someone loses a life, they must recover it by gathering potions. Alternatively, one can gain experience (XP) bars by killing weak foes via bullets shot and capturing them. One extra life is awarded once the XP meter is full. Such an endeavor should, however, consider the remaining ammunition and the 90-second time restriction. These restrictions prevent players from exploiting such tactical strategies purely for amusement while ignoring the main objective of the game, which is to keep compound formation in the player's memory.

According to [35], the game can procedurally generate up to 486,000 playable stages. Each stage consists of a combination of learning material and a game map. This vast number of game stages allows players to experience different challenges categorized into three difficulty levels. The game map data, the player's actions, and achievements are recorded during game sessions. These data together are called in-game data.

B. Proposed Methodology

This research follows some procedures shown in Figure 2. The first step collects datasets from the Chem Dungeon game sessions using the procedures shown in Figure 3. A survey was conducted to allow participants to follow the data collection steps. Each participant played at least ten cycles of data collection. This means that each cycle will produce a sample comprised of in-game data and a label. The label (a.k.a Memorization Performance/MP) is the score difference between pre (M_0) and post-game (M_1) questionnaires about the learning material presented in the game stage. Given that each response to the questionnaire is a binary value, there will be four possible MP categories shown in Table 1.

From Table 1, there are only three categories used. First, if the players score 0 between the pre and post-game, the sample is categorized as MP0. It represents a player who needs to memorize new knowledge. Second, MP1 is the label when the difference between pre and post-game is 1. It means a recognition (successful Memorization) of the new knowledge. The third category is MP2, when players can recall knowledge they already know. The fourth label does not use a negative difference between pre and post-game, indicating a decrease in memory. These results occurred because a player arbitrarily responded to the pre or post-game questionnaires, so the sample was categorized as

an outlier [36]. This method is simple; however, the risk of irrelevant gaming action or arbitrary responses to the questionnaire is possible. Therefore, we must preprocess the resulting in-game data before the modeling stage.
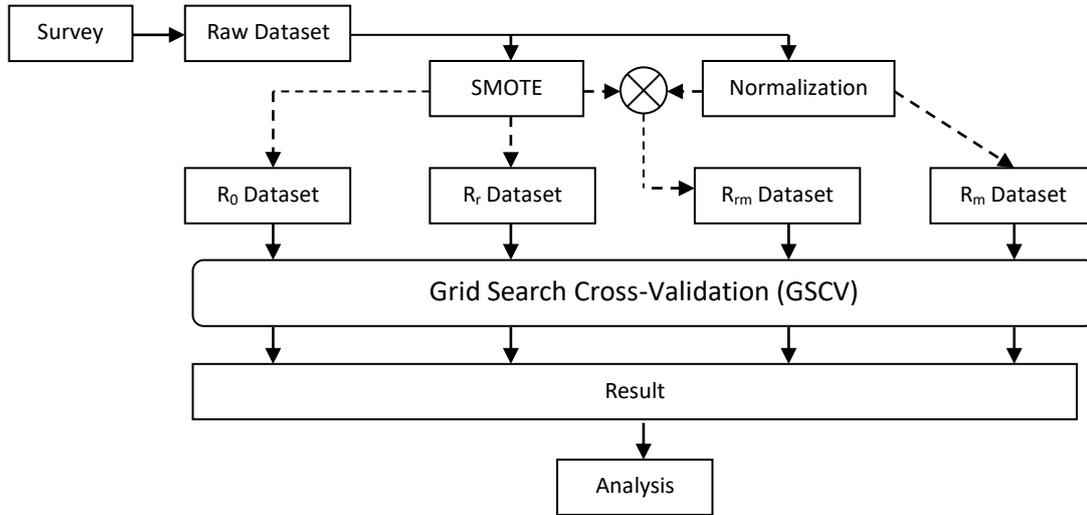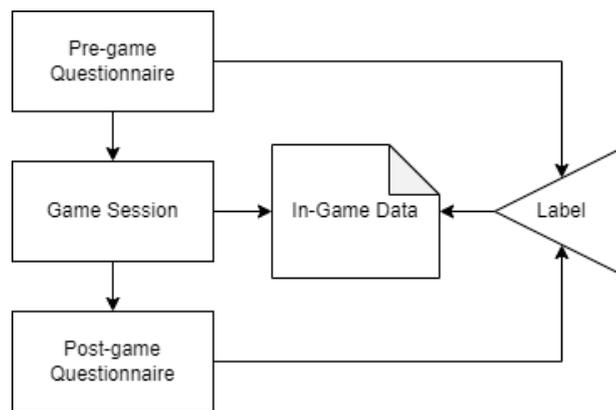

Fig 2. Research procedures

Fig 3. Data collection procedures

Table 1. Memorization Performance

	$M_1 = 0$	$M_1 = 1$
$M_0 = 0$	MP0	MP1
$M_0 = 1$	N/A	MP2

Once the dataset is collected, the next step is the preprocessing stage. Preprocessing aims to identify outliers by filtering out samples with a negative label or missing values. This results in a clean Raw Dataset, the first dataset S_0. The following preprocessing are resampling for imbalanced dataset and min-max normalization, or their combination. As such, preprocessing yields three more datasets: Smoted Dataset S_r, Normalized Dataset S_m, and SmotedNormalized Dataset S_{rm}. Each dataset is split into a 70% training set (i.e., R_0, R_r, R_m, R_{rm}) and a 30% test set (i.e. T_0, T_r, T_m, T_{rm}). Each dataset (i.e., R_0, R_r, R_m, R_{rm}) will be used to construct an optimized classifier using the Random Forest Algorithm. Random Forest can generalize well given the high-dimensional dataset with higher accuracy than other algorithms [34][37]. Research in [34] shows that the Random Forest well classifies behavior-related data. The in-game data falls into this category. The first optimization targets the Random Forest parameter using the Grid Search Cross-Validation (GSCV). The second optimization is to evaluate whether preprocessing affects the classification result.

The following are the GSCV configurations for the Random Forest Classifier using the training dataset:

- Parameter grid is n_estimators = {2,3,…,201},
- 5-fold cross-validation (considering the imbalanced dataset),
- Scorer = weighted F1-score (considering the imbalanced dataset)

Then, we trained the Random Forest (RF) using the training set over the best value of n_estimators; we call these values n_searched. We measured the RF performance using the F1-score because it accommodates precision and recall scores. Here is the formula for F1-score = 2*Precision*Recall/(Precision+Recall). The resulting list of RFs trained using training sets is called RF_searched. However, the best RF should not solely be confirmed from GSCV peak performance. So, we delved deeper into the GSCV statistics to see the overall picture of RF_searched based on the average mean scores (A_mean) and the average standard deviation of the scores (A_std).

We observed the region of interest in both graphs. From there, we decided the best RF from the one closest to the A_mean and the lowest A_std. Then, we retrained the Random Forest classifier using the best n_estimators using the training sets (i.e., R_0, R_r, R_m, and R_{rm}). We used the test sets (i.e., T_0, T_r, T_m, and T_{rm}) to evaluate the performances of each classifier. The goal is to reproduce the training stage from the GSCV. By testing each optimal Random Forest classifier using the test set (i.e., T_0, T_r, T_m, T_{rm}), one can compare the effect of normalization, balancing, balancing-normalization, and n_estimators for the classification. In the final stage, we tested the performance difference between optimized classifiers; we used McNemar's Test [38].

III. Result and Discussion

From the survey, according to Figure 3, we collected 540 samples of in-game data labeled with MP_0, MP_1, and MP_2 (this is the raw dataset S_0). We distributed 90, 219, and 231 samples to MP_0, MP_1, and MP_2. Each sample contains 30 independent variables of mixed types. Subsequently, we generated S_r consisting of 744 samples equally distributed via SMOTE from the original dataset (S_0). Next, we developed S_m consisting of 540 min-max normalized samples from the original dataset (S_0). Subsequently, the dataset S_{rm} (744 samples) was generated from S_m by resampling the normalized dataset. S0, Sr, Sm, and Srm datasets were split into a 70% training set (i.e., R_0, R_r, R_m, and R_{rm}) and a 30% test set (i.e., T_0, T_r, T_m, and T_{rm}).

The optimization stage using GSCV identifies four classifiers: C_0, C_r, C_m, and C_{rm}, best constructed using 89, 31, 89, and 196 trees, respectively. See Figure 4 for comparing these four classifiers that predict a player's Memorization given the test set (i.e., T_0, T_r, T_m, and T_{rm}). All classifiers successfully predicted the Memorization of the players via in-game data with at least 80% confidence. However, we can see that the C_r was the better, with overall scores of ~86%. From this graph, we can see that the balanced dataset is slightly better than the imbalanced dataset. We use these performance rates to optimize the classifiers via n_estimators.

We analyze further the GSCV results in the improvements made with various values of n_estimators. Figure 5 to Figure 8 show the comprehensive results of each Random Forest classifier when the GSCV searched the best n_estimators or the number of trees. There are four line graphs:

- The blue solid line represents the mean f1 scores of the RF_searched (using the left horizontal axis),
- and the blue dotted line represents the average of mean f1 scores (using the left horizontal axis). We denote this as avg_f1,
- The red solid line represents the standard deviation of f1 scores between cross-validated predictors of the RF_searched (using the right horizontal axis),
- the red dashed line represents the average of std deviation of f1 scores between cross-validated predictors of the RF_searched (using the right horizontal axis). We denote this as avg_std.
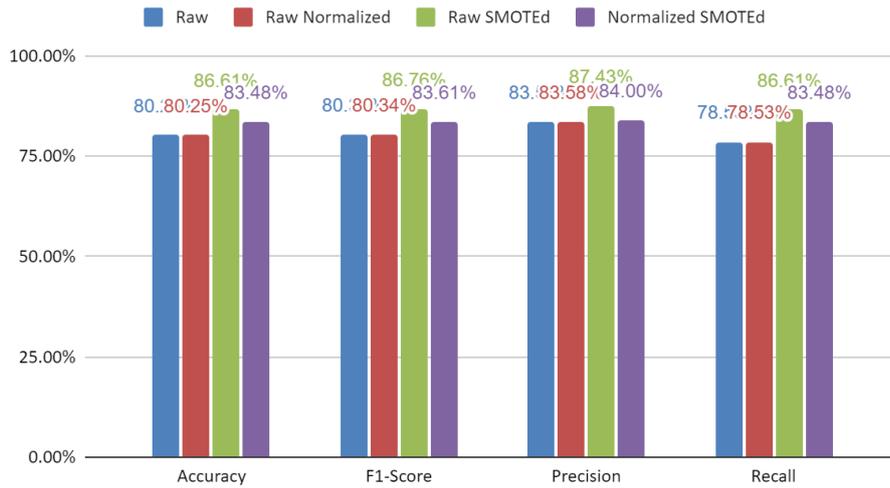
Fig 4. Performance comparison between classifiers using BEST n_estimators

In addition, there are also two rectangles (transparent blue and transparent red) representing the regions of interest (ROI_blue and ROI_red) we observed regarding the candidate value of n_estimators that optimizes the scores. The right bound of each rectangle is set based on the best n_estimators found from the GSCV stage. Meanwhile, the left bound is set towards the lowest possible n_estimators value where the f1 score is greater or equal to the blue dotted line. The rules to choose the optimal n_estimators are:

- Choose the value of n_estimators (in the x-axis) from the left most of the ROI_blue. We denote FSx as the mean f1 score from the selected RF using the n_estimators value.
- Choose the current value of n_estimators if the FSx* of the next n_estimators is less than or equal to the current FSx.
- IF some neighboring n_estimators have the same FSx, then choose the n_estimators which has the smallest value of avg_std.
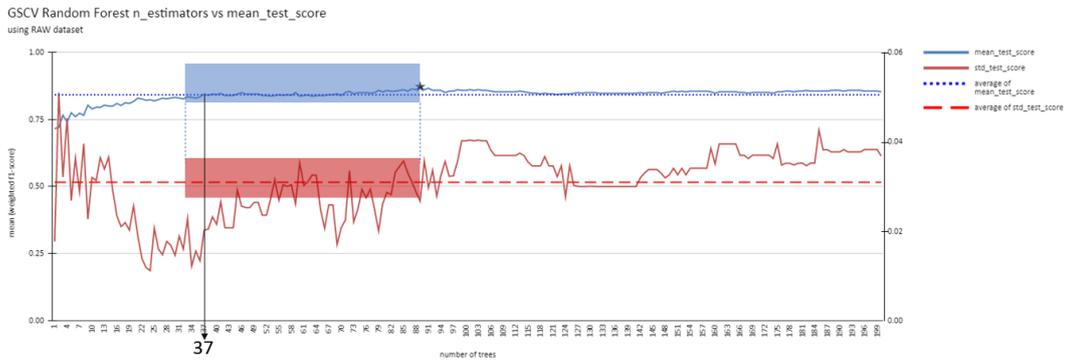

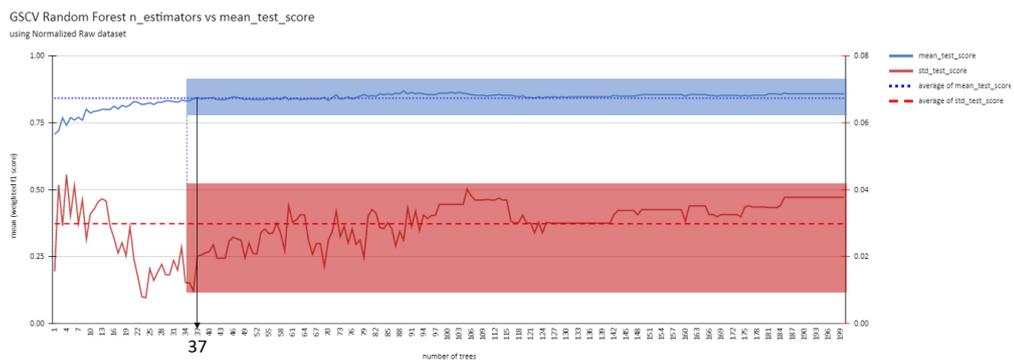Fig 5. GSCV random forest using raw training set

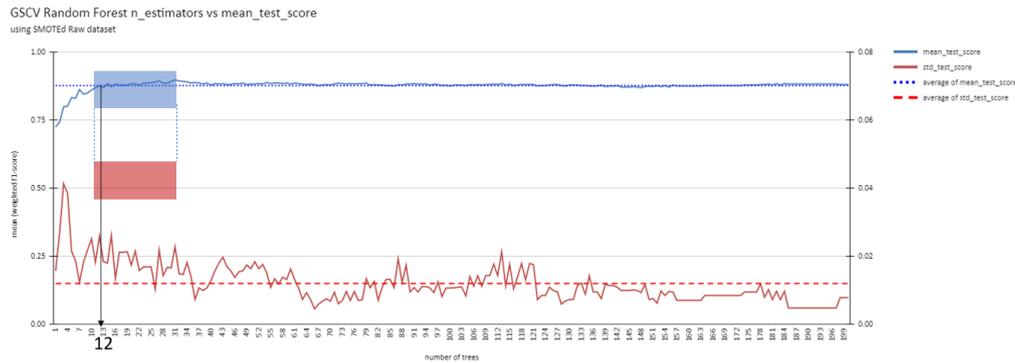Fig 6. GSCV random forest using normalised training set

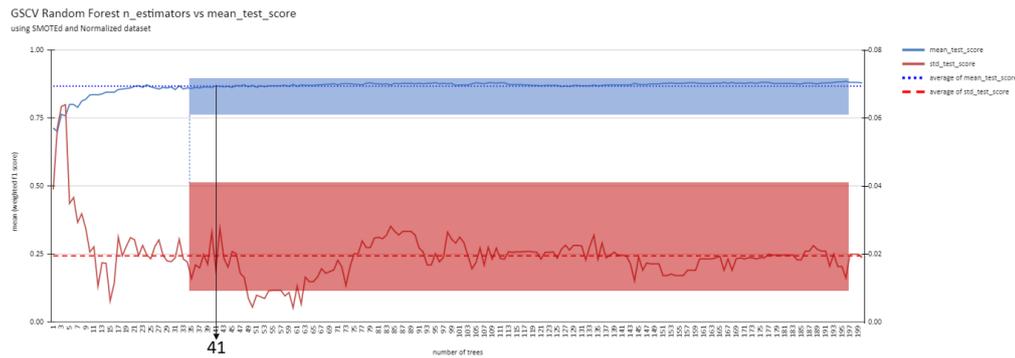Fig 7. GSCV random forest using SMOTEd raw training set

Fig 8. GSCV random forest using SMOTEd normalized training set

These graphs show that the lowest F1-scores were around 0.7 - 0.73 and quickly stabilized between 0.86 - 0.89 of f1 scores. This indicates that the Random Forest classifiers were effective under the 5-fold cross-validation using the training sets. Based on the above rules, we identified the values of n_estimators for C_0, C_r, C_m, and C_{rm} using 37, 12, 37, and 41 trees, respectively (Table 2).

Table 2. Comparison of best and optimal RF based on n_estimators

GSCV Random FOrest	n_estimators		scores			
	best GSCV	optimal GSCV	best score	optimal score	best std	optimal std
Raw	89	37	0.8718033199	0.8447963115	0.02667657109	0.02016542199
Raw SMOTEd	31	12	0.8757689417	0.8748399185	0.01188986414	0.02573413102
Raw normalized	89	37	0.8689533825	0.8447963115	0.02490655104	0.02016542199
SMOTEd normalized	196	41	0.8844496674	0.8676858415	0.01288923776	0.01528713981

Next, we retrained the C_0, C_r, C_m, and C_{rm} using 37, 12, 37, and 41 trees, respectively, using the training sets. Comparing classifiers that used OPTIMAL value of n_estimators can be seen in Figure 9. This graph shows that the Random Forest maintained the prediction performances when using significantly fewer trees. The SMOTEd datasets make the classifiers slightly more steady than the imbalanced datasets (S_0 and S_m). In addition, the classifiers fitted using the SMOTEd-normalized dataset maintained their performances using only 41 trees, compared to the 196 trees initially found in GSCV.

Based on these performances, we ran McNemar's Test to see if there were any significant differences between classifiers using BEST and OPTIMAL n_estimators. As a result, all p-values were at least 0.05. This indicates that these optimal classifiers were similar. It means that the Random Forest algorithm is a robust classifier and optimizable via the total decision trees used to predict the memorization performance of Chem Fight players.

These experiments prove our confidence in using Random Forest as the classifier to predict the player's Memorization. The raw dataset, which was not resampled nor normalized, can be classified well. Upon the optimization, experiment results show that resampling the dataset using SMOTE can improve the performance of the Random Forest to at least 4% higher. We proved that the

GSCV method can slightly optimize the performance of Random Forest. Because we only optimize the Random Forest's n_trees, while there are more parameters that are optimizable such as max_depth, min_samples_split, min_samples_leaf, min_weight_fraction_leaf, max_features= {"sqrt", "log2", None}, max_leaf_nodes.
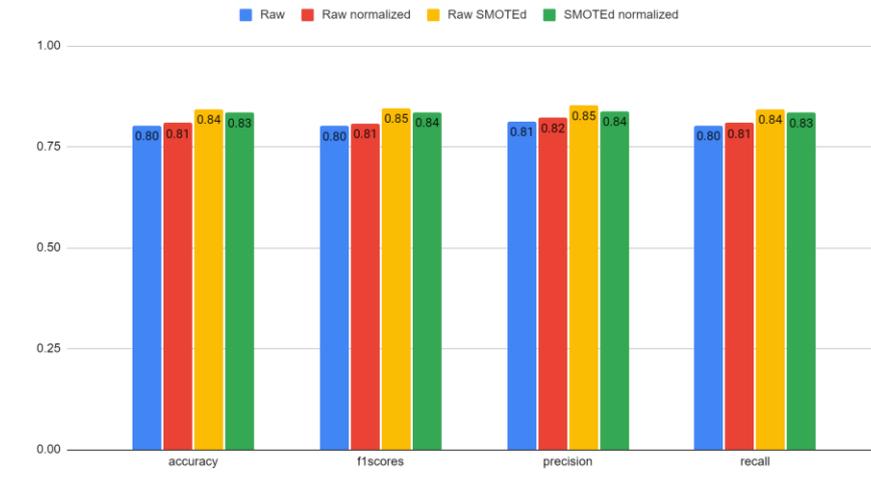

Fig 9. Performance comparison between classifiers using OPTIMAL n_estimators

IV. Conclusion

Assessing a game education player's Memorization is non-intrusively practical using in-game data. Our approach applies a data mining classification technique using the Random Forest (RF) algorithm. We experimented using variants of the dataset to train the RF. Since RF is a complex classifier, we used a Grid Search Cross-Validation (GSCV) technique to identify and picture the development of classifiers based on a vector or n_estimators. Our approach has successfully optimized the classifiers that use at least half the total trees inside the RF. The classifiers predict the player's Memorization with around 80% accuracy using the imbalanced dataset and using 37 decision trees (optimal). The classifiers performed better (~86% accuracy) when fitted with a balanced dataset. We also found that the most effective optimization occurred when the classifier used the balanced and normalized dataset. Based on our experiments, the n_estimators found by the GSCV are based on the peak performance of the classifiers. Our observation identified that the classifiers maintained their performance at least using half the value of n_estimators found by the GSCV.

Our experiments demonstrated Random Forest's suitability for predicting player Memorization without data preprocessing. However, applying SMOTE to the dataset boosted Random Forest's performance by at least 4%, and GSCV showed slight optimization potential. Further optimization possibilities include parameters like max_depth, min_samples_split, and more. Our approach optimized the Random Forest based on n_estimators using GSCV. However, when considering multiple hyperparameters in optimizing the classifiers, applying GSCV may become more complex. Hence, we suggest using a more sophisticated search algorithm, such as Genetic Algorithm Search Cross-Validation. In addition, a mechanism in the search algorithm to stop earlier whenever the performance of the classifiers enters the convergence state.

We are confident that other researchers can replicate our procedure to determine the optimal classification of players in other games. However, we know that the selection of in-game actions can bias the classification model—for instance, too few or arbitrary in-game actions. For now, these in-game actions are human-observed ones. In contrast, low-level in-game activities, such as player positions and time-based events, are too noisy to be classified using RF and are not interpretable for humans. Hence, a Neural Network or the Deep Neural Network is a potential candidate for this classification problem. Given that the low-level in-game is preferable, we can identify them via behavior recognition computationally.

Declarations

Author contribution

All authors contributed equally as the main contributor of this paper. All authors read and approved the final paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare no known conflict of financial interest or personal relationships that could have appeared to influence the work reported in this paper.

Additional information

Reprints and permission information are available at http://journal2.um.ac.id/index.php/keds.

Publisher's Note: Department of Electrical Engineering and Informatics - Universitas Negeri Malang remains neutral with regard to jurisdictional claims and institutional affiliations.

References

[1] J. P. Gee, "What video games have to teach us about learning and literacy," Comput. Entertain., vol. 1, no. 1, pp. 20–20, Oct. 2003.

[2] J. J. C. U. Lee and J. C. U. Hammer, "Gamification in Education: What, How, Why Bother?," Acad. Exch. Q., vol. 15, no. 2, pp. 1–5, 2011.

[3] K. Squire and H. Jenkins, "Harnessing the power of games in education," Insight, vol. 3, no. 1, pp. 5–33, 2003.

[4] V. Shute et al., "Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity and learning supports in educational games," J. Comput. Assist. Learn., vol. 37, no. 1, pp. 127–141, Feb. 2021.

[5] K. Mayfield et al., "Designing a Molecular Biology Serious Educational Game," in Proceedings of the 2019 ACM Southeast Conference, Apr. 2019, pp. 210–213.

[6] J. P. Gee, "Good video games and good learning," Univ. Wisconsin–Madison. Recuper., 2009.

[7] S. Çiftci, "Trends of Serious Games Research from 2007 to 2017: A Bibliometric Analysis.," J. Educ. Train. Stud., vol. 6, no. 2, pp. 18–27, 2018.

[8] O. Irmade, Suwarno, and N. Anisa, "Research Trends of Serious Games: Bibliometric Analysis," J. Phys. Conf. Ser., vol. 1842, no. 1, p. 012036, Mar. 2021.

[9] N. Kara, "Bibliometric and Content Analysis of Research Trends on the Use of Serious Games to Assist People with Disabilities," J. Comput. Educ. Res., vol. 9, no. 17, pp. 278–299, Apr. 2021.

[10] S. MEŞE and C. MEŞE, "Research Trends on Digital Games and Gamification in Nursing Education," J. Comput. Educ. Res., vol. 10, no. 20, pp. 734–750, Dec. 2022.

[11] S. P. Smith, K. Blackmore, and K. Nesbitt, "A Meta-Analysis of Data Collection in Serious Games Research," in Serious Games Analytics, Cham: Springer International Publishing, 2015, pp. 31–55.

[12] R. Aylett, M. Vala, P. Sequeira, and A. Paiva, "FearNot! – An Emergent Narrative Approach to Virtual Dramas for Anti-bullying Education," in Virtual Storytelling. Using Virtual Reality Technologies for Storytelling, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 202–205.

[13] M. D. Kickmeier-Rust, C. Hockemeyer, D. Albert, and T. Augustin, "Micro Adaptive, Non-invasive Knowledge Assessment in Educational Games," in 2008 Second IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning, 2008, pp. 135–137.

[14] M. D. Kickmeier-Rust, An alien's guide to multi-adaptive educational computer games. Informing Science, 2012.

[15] M. D. Kickmeier-Rust, C. M. Steiner, and D. Albert, "Non-invasive Assessment and Adaptive Interventions in Learning Games," in 2009 International Conference on Intelligent Networking and Collaborative Systems, Nov. 2009, pp. 301–305.

[16] B. Magerko, C. Heeter, J. Fitzgerald, and B. Medler, "Intelligent adaptation of digital game-based learning," in Proceedings of the 2008 Conference on Future Play: Research, Play, Share, Nov. 2008, pp. 200–203.

[17] F. Bellotti, R. Berta, A. De Gloria, and L. Primavera, "Adaptive Experience Engine for Serious Games," IEEE Trans. Comput. Intell. AI Games, vol. 1, no. 4, pp. 264–280, Dec. 2009.

[18] A. Plotnikov et al., "Measuring enjoyment in games through electroencephalogram (eeg) signal analysis," in Proceedings of the 6th European Conference on Games-Based Learning (ECGBL 2012), 2012, pp. 393–400.

[19] os, K. Lee, P. Moreno-Ger, and R. Berta, "Assessment in and of Serious Games: An Overview," Adv. Human-Computer Interact., vol. 2013, pp. 1–11, 2013.

[20] D. Ismailović, J. Haladjian, B. Köhler, D. Pagano, and B. Brügge, "Adaptive serious game development," in 2012 Second International Workshop on Games and Software Engineering: Realizing User Engagement with Game Engineering Techniques (GAS), 2012, pp. 23–26.

[21] B. Magerko, B. Stensrud, and L. Holt, "Bringing the schoolhouse inside the box-a tool for engaging, individualized training," 2006.

[22] J. Rowe and J. Lester, "Modeling User Knowledge with Dynamic Bayesian Networks in Interactive Narrative Environments," Proc. AAAI Conf. Artif. Intell. Interact. Digit. Entertain., vol. 6, no. 1, pp. 57–62, Oct. 2010.

[23] T. Hainey et al., "Serious games as innovative formative assessment tools for programming in Higher Education," 2022.

[24] P. F. Lazarsfeld, "The language of social research: A reader in the methodology of social research," 1964.

[25] F.-L. Fu, R.-C. Su, and S.-C. Yu, "EGameFlow: A scale to measure learners' enjoyment of e-learning games," Comput. Educ., vol. 52, no. 1, pp. 101–112, Jan. 2009.

[26] R. Lopes and R. Bidarra, "Adaptivity Challenges in Games and Simulations: A Survey," IEEE Trans. Comput. Intell. AI Games, vol. 3, no. 2, pp. 85–99, Jun. 2011.

[27] C. Alonso-Fernández, M. Freire, I. Martínez-Ortiz, and B. Fernández-Manjón, "Improving evidence-based assessment of players using serious games," Telemat. Informatics, vol. 60, p. 101583, Jul. 2021.

[28] Devottam Gaurav, Yash Kaushik, Santhoshi Supraja, Manav Yadav, M P Gupta, and Manmohan Chaturvedi, "Empirical Study of Adaptive Serious Games in Enhancing Learning Outcome," Int. J. Serious Games, vol. 9, no. 2, pp. 27–42, May 2022.

[29] A. S. Osman, "Data Mining Techniques: Review. 2 (1), 1–4." 2019.

[30] J. Han, M. Kamber, and J. Pei, "Data mining, southeast asia edition: Concepts and techniques. 2006." Morgan kaufmann.

[31] J. A. Caballero-Hernández, M. Palomo-Duarte, J. M. Dodero, and D. Gaševic, "Supporting Skill Assessment in Learning Experiences Based on Serious Games Through Process Mining Techniques," Int. J. Interact. Multimed. Artif. Intell., vol. In Press, no. In Press, p. 1, 2023.

[32] S. Abbasi and H. Kazi, "Stealth assessment in serious games to improve OO learning outcomes," in 2019 International Conference on Advances in the Emerging Computing Technologies (AECT), Feb. 2020, pp. 1–5.

[33] M. Manske and C. Conati, "Modelling Learning in an Educational Game," in AIED, 2005, vol. 2005, pp. 411–418.

[34] Z. Feng, Y. Shi, D. Zhou, and L. Mo, "Research on Human Activity Recognition Based on Random Forest Classifier," in 2023 IEEE International Conference on Control, Electronics and Computer Technology (ICCECT), Apr. 2023, pp. 1507–1513.

[35] H. A. Rosyid, M. Palmerlee, and K. Chen, "Deploying learning materials to game content for serious education game development: A case study," Entertain. Comput., vol. 26, pp. 1–9, May 2018.

[36] L. Bao, "Dynamic models of learning and education measurement," arXiv Prepr. arXiv0710.1375, 2007.

[37] A. Statnikov, L. Wang, and C. F. Aliferis, "A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification," BMC Bioinformatics, vol. 9, no. 1, p. 319, Dec. 2008.

[38] T. G. Dietterich, "Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms," Neural Comput., vol. 10, no. 7, pp. 1895–1923, Oct. 1998.