2022 IEEE World Artificial Intelligence & Internet of Things Congress (AIIoT) | 978-1-6654-8453-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/AIIoT54504.2022.9817177

Clustering and Classification Models For Student's Grit Detection in E-Learning

Renato R. Maaliw III
College of Engineering
Southern Luzon State University
Lucban, Quezon, Philippines
rmaaliw@slsu.edu.ph

Karen Anne C. Quing
College of Arts & Sciences
Southern Luzon State University
Lucban, Quezon, Philippines
kquing@slsu.edu.ph

Julie Ann B. Susa
College of Engineering
Southern Luzon State University
Lucban, Quezon, Philippines
jsusa@slsu.edu.ph

Jed Frank S. Marqueses
College of Arts & Sciences
Southern Luzon State University
Lucban, Quezon, Philippines
jmarqueses@slsu.edu.ph

Ace C. Lagman
Information Technology Dept.
FEU Institute of Technology
Manila, Philippines
aclagman@feutech.edu.ph

Rossana T. Adao
College of Computer Studies
FEU Institute of Technology
Manila, Philippines
radao@fit.edu.ph

Ma. Corazon Fernando - Raguro
Information Technology Dept.
FEU Institute of Technology
Manila, Philippines
mgfernando@feutech.edu.ph

Ranie B. Canlas
College of Computer Studies
Don Honorio Ventura State University
Bacolor, Pampanga, Philippines
raniecanlas@dhvsu.edu.ph

***Abstract*—Grit plays a crucial role in determining high individual success more than intellectual talent alone. However, there is no existing literature that ventured into the trait identification in an e-learning environment. This study presents a comprehensive computational-driven strategy for detecting a learner's grit using machine learning. Empirical results show that DBSCAN and Random Forest models produce average accurate prediction consistency of 92.67% against the questionnaire method. Knowledge interpretation using feature importance and association mining quantifies perseverance and sustained interest as the most pressing component of grit. Correlational analysis reveals that grit has a weak connection with course grades (short-term goal) but demonstrates a strong positive association with professional achievement (long-term goal) and maturation. Collectively, our findings substantiate that breakthrough accomplishment is contingent not solely on cognitive ability but on constant interests and resilience.**

Keywords—classification, density-based clustering, dynamic time warping, grit scale, non-cognitive traits, prediction

I. Introduction

The significance of intellectual capacity is well documented in academic and professional domains, but we understood very little about other equally critical characteristics that foresee success. While various demographic features have a role in both contexts, many aspects influence achievements, including personality and behavioral factors. Studies have proven that learners' cognitive ability is insufficient [1]. They must surmount hurdles, be dedicated and disciplined, and maintain high motivation to achieve goals while remaining resilient in any environment or situation [2] [3].

Along with the factors above, researcher [4] presented grit as an essential trait for success in her seminal work. Grit is an intangible individual quality of perseverance, willingness, and determination to accomplish long-term aspirations despite setbacks. It correlates positively with intrinsic motivation to pursue breakthrough objectives under challenging circumstances. Gritty people continually push to give their best without regard for constraints, boundaries, or thresholds – instead, they attempt to learn, comprehend and excel to the maximum extent possible [5]. Likewise, it is essential to understand that grit is not innate but dynamic; it can be identified and cultivated. However, one needs to adopt a growth mindset to build grit by recognizing that intelligence and talent can be fostered through hard work and the willingness to explore new approaches as an opportunity to learn and progress.

Existing literature reveals that high grit levels consistently predicted individual success more accurately than pure intelligence alone regarding educational and non-academic outcomes, as evidenced by [6 – 8]. For example, it predicted spelling bee performance, graduation rates, and higher scholastic ratings. These conclusions are confirmed by the report of [9] that examines grit in graduate students' accomplishments in online courses. Results show a substantial link between grit and grade point average (GPA), study time duration, and age. Furthermore, it was attributed to excellent retention among military cadets, junior and senior high schools, and long-term employment [10].

For quantifying the characteristic, a 12-item grit scale, a questionnaire designed by the author [11], assesses personality attributes with scores ranging from 1 to 5 (low to high grit). Over the past decades, the scale has become a standard adopted widely by practitioners in numerous forms of achievement research. However, published articles laid out the problems with the questionnaire's construct, casting doubts on its claims. It ranges widely from statistical to methodological errors [12 – 14]. Conversely, surveys are prone to inaccuracies due to social desirability and frame of reference bias [15], different interpretations of questions, thoughtless responses [16], and survey fatigue [17].

While considerable efforts employed grit as a tenet to predict success, no existing studies glanced into the prospect of identifying the trait through learners' behaviors in an e-learning environment using computational-driven methods. If done right, this analysis can help establish machine learning models to identify genuine grit levels automatically compared to simple questionnaires. Its potential discoveries have applications in computational psychology [18], personalized learning [19], and recommender systems, to name a few.

Grit analysis in a virtual learning environment (VLE) is still in its inception. In response to the identified gaps, we conducted the first experiment of its kind to investigate the

Commission on Higher Education (CHED) – Institutional Development and Innovation Grants (IDIG)

plausibility of detecting students' grit in remote learning. E-learning entails a constructivist approach, requiring learners to work independently to develop ideas, solve issues, and acquire knowledge. Individuals who cannot self-regulate (an element of grit) may underperform in online classes and soon lose the desire to succeed. The study's relevance cannot be overstated, as e-learning, pandemic or not, has become increasingly prevalent and an integral part of all higher educational institutions (HEIs) worldwide.

Our research's primary contribution is the generated machine learning models for identifying grit in e-learning to advance the discourse meaningfully by interpreting the association between grit and achievements in academic and professional settings. This initiative will aid educators at all levels by utilizing innovative data analytics to ascertain student characteristics and deliver interventions to help them build non-cognitive abilities.

II. METHODOLOGY

A. Online Course Description and Data Sources

The research was inspired by a Moodle course in Programming 1 (C++), an online complementary learning resource to traditional face-to-face classes for sophomore tertiary students. It comprises 86 well-crafted learning objects divided into six chapters with simulations, self-assessment exams, quizzes, videos, and other instructional materials to help students learn to code. Moreover, it has 562,174 recorded interaction logs that span four years (2014 to 2017) – the most dynamic course.

We purposively selected the 500 most active accounts based on equal standard distributions. Aside from their e-learning log data, we also collected their grit scale questionnaire results, final course grades, and current job levels (professional achievement) for correlation purposes from the college's record section and tracer study surveys. We also removed students' data that dropped out of the course to prevent outliers.

B. Grit Related Learner Attributes in E-learning

We extracted pertinent features from the learning management systems (LMS) database to create the initial dataset. As described in the literature, gritty people are motivated, persistent, self-disciplined, meticulous, resilient, and have sustained interest. To gauge self-motivation, we tracked the number of instances a learner attempted to reach a passing score by conducting self-assessment examinations. These are not graded exams. Such undertaking demands a superior degree of intrinsic drive.

For self-discipline traits, we mapped each student's hours spent in the course outside the authorized class schedule, as this demonstrates self-regulation. Detailed-oriented or diligent persons are more likely to complete mandatory tasks on time. Thus, the collection of data regarding students' submissions was necessary. We then tallied learners' attempts to finish and provide correct solutions to complex programming simulations to measure perseverance. Complex coding challenges require resiliency in the face of repeated failures [20]. Finally, we analyze a student's consistency of logins to the course to quantify sustained interests.

C. Feature Mapping for Clustering

Table 1 depicts the mapping of students' extracted data from the LMS database according to the characterization of grit detailed in the previous section.

TABLE 1
STUDENTS' GRIT ATTRIBUTES IN E-LEARNING

Attribute	Relevant Behavior (Variable Name)	Value/Type	Sample/ Encoding
Self-Motivation	Number of completed self-assessment tests attempts until reaching a passing score (self-motivation)	Total attempts/ integer	e.g. 10, 12
Self-discipline	Number of hours spent in the online course beyond the scheduled class hours (self-discipline)	Total hours/ float	e.g. 25.3, 38.7
Diligence	Number of submitted mandatory tasks on time (diligence)	Total submissions/ integer	e.g. 3,5,7
Perseverance	Number of completed attempts until finding a correct solution to complex programming problems (perseverance)	Total attempts/ integer	e.g. 12, 14
Sustained interest	Consistency of logins in the online course beyond the scheduled class hours per week (interest)	Consistency category/ integer	2 = consistent 1 = inconsistent

D. Access Consistency Calculation

A different modeling technique was necessary to estimate the sustained interest attribute's values. First, we applied dynamic time warping (DTW), a technique for comparing two temporal sequences [21]. It is designed to generate distance metrics between two input time series. Typically, the resemblance or disparities is determined by transforming the data into vectors and computing the Euclidean distance (*d*) (when $P = 2$) between vector space points as described by Equation 1:

$$d(X,Y) = \left(\sum_{i=1}^{n} |x_i - y_i|^P\right)^{\frac{1}{P}} \tag{1}$$

Where X and Y are two different time series data of the same type, and x_i and y_i are the vector points for comparison. If two sequences are equal, their distance is zero [22], while a distance value of > 1.20 indicates dissimilarity based on the simulated T-test p-value (< 0.05). We set a consistency baseline based on the average login hours of all learners for 18 weeks, equivalent to a semester. According to our data, there are instances when an individual surpassed the weekly login baselines. To avoid penalizing them for exceeding the threshold, we set their login hours equal to the derived maximum benchmark. In plain words, if a specific weekly average baseline is equivalent to 3 hours and a student has logged in for 4 hours (beyond the baseline), then student login hours are reverted to 3 (maximum). Figures 1(a) and 1(b) show a DTW excerpt for determining consistency, including the calculated distances.

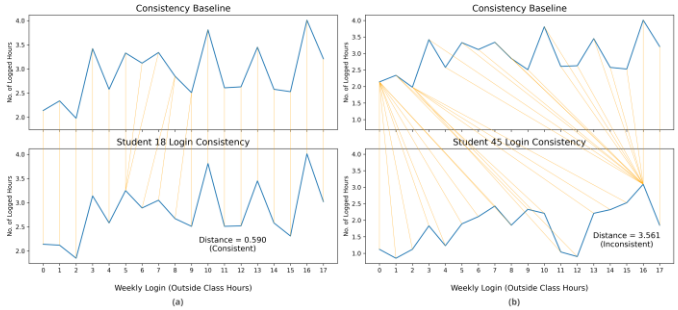

Fig. 1. Consistent logins (a) vs. Inconsistent logins (b) based on DTW

E. Grit Level Detection Through Clustering

Although we collected their grit levels through grit scale questionnaires, we assessed the trait using a density-based spatial clustering (DBSCAN) [23]. It groups together points that are close to each other based on a Euclidean distance metric and the minimum number of points to form a dense region. Moreover, the method efficiently diminishes noise and catches outliers offering better results than other clustering methods [24]. First, we applied a *min-max* normalization to transform the original data into well-distributed ranges between 0 and 1. This procedure reduces the impact of larger variables over smaller ones, avoiding convoluted clusters. Equation 2 displays the formula:

$$d' = \frac{d - d_{min}}{d_{max} - d_{min}} \quad (2)$$

Where d' is the scaled data, d_{max} and d_{min} are the maximum and minimum feature values. Second, we empirically tested all two-dimensional combinations of variables corresponding to Table 1 and selected the highly separable clusters based on silhouette scores [25] shown in Figure 2.

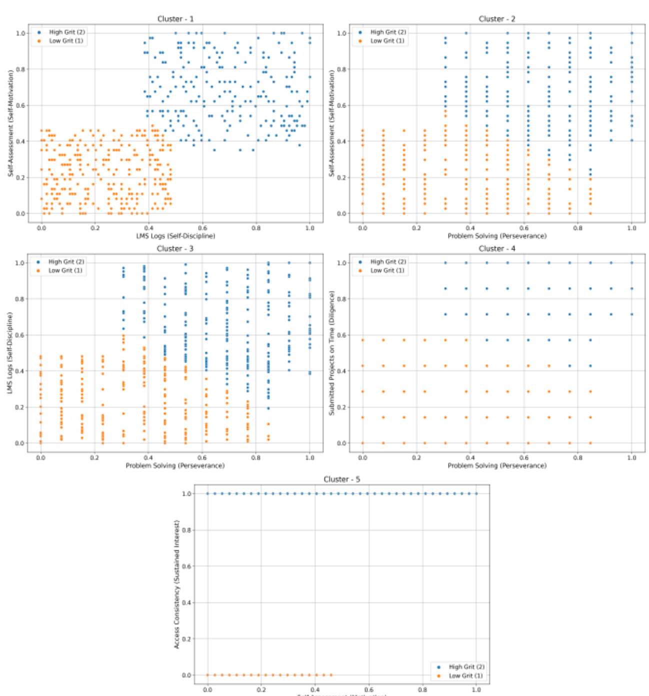

Fig. 2. Five highly separable clusters for grit level identification based on silhouette scores (C1 = 0.8, C2, = 0.7, C3 = 0.7, C4 = 0.6, C5 = 0.6)

Finally, the final grit levels of each individual were identified through majority votes (similar to ensemble modeling's hard voting mechanics) from each of the five clusters. This approach enables the estimation to be more robust. Table 2 shows an excerpt of the process.

TABLE 2
EXCERPT OF CLUSTER VOTING

Student ID	Clusters					Grit Level/ (Encoding)
	Cluster-1	Cluster-2	Cluster-3	Cluster-4	Cluster-5	
18	High Grit	High Grit	Low Grit	High Grit	High Grit	High Grit/ (2)
45	Low Grit	Low Grit	Low Grit	High Grit	Low Grit	Low Grit/ (1)
121	High Grit	Low Grit	High Grit	High Grit	High Grit	High Grit/ (2)

F. Grit Level Prediction Through Classification

Our study further pushed the boundaries by using classification to validate the reliability of the clustering results. This validation technique [26] can supplement a virtual learning environment's grasp on grit identification. We partitioned the dataset (70/30 training & testing split) with features corresponding to Table 1 and assigned class labels of 'High Grit' and 'Low Grit'; encoded as 2 and 1 (see Figure 2 and Table 2). Next, we implemented 10-fold cross-validation to evaluate the accuracies of various supervised learning models, including J48 Decision Tree, Logistic Regression (LogReg), Artificial Neural Network (ANN), K-Nearest Neighbors (KNN), Random Forests (RF), Support Vector Machines (SVM), and Naïve Bayes (NB). This strategy enables each classifier to learn on equal proportions of categories to eliminate bias. The optimum hyperparameters were uncovered using a grid search algorithm [27] by automatically comparing adjusted accuracy variations for each model versus manual tuning.

G. Dimensionality Reduction

High-dimensional data has the inevitable effect of increasing prediction complexity while decreasing generalizability. As a remedy, variable elimination is needed to construct simplified datasets. The compromise is to lose inconsequential information in exchange for model simplicity. We utilized principal component analysis (PCA), an unsupervised feature selection technique. According to standards, an explained variance ratio of 70% is typical for conserving certainty [28]. Figure 3 represents that PC-1 (81.53%) and PC-2 (9.23%) contained a total of 90.76% variations.

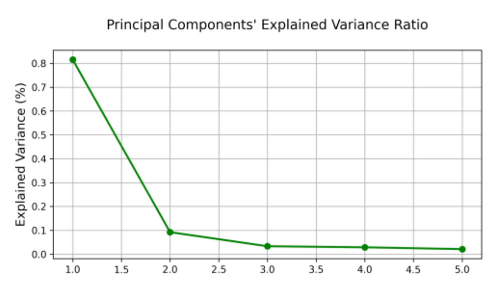

Fig. 3. Principal component's explained variance ratio

Table 3's feature ranking stressed that perseverance (0.434), sustained interest (0.415), self-motivation (0.392), and self-discipline (0.383) are the primary contributors to each component. Diligence (0.123 & 0.121), however, ranked the least. Thus, we excluded the feature and made sure to balance the distribution for each class by under sampling with 232 each ('high grit' & 'low grit, N = 464). It enables the trained models to prevent poor results and improve generalization.

TABLE 3
FEATURE RANKING BY COMPONENTS

Variables	PC - 1	PC - 2	PC - 3	PC - 4
Perseverance	**0.434**	0.113	0.028	0.018
Sustained interest	**0.415**	0.121	0.079	0.022
Self-motivation	**0.392**	0.125	0.132	0.023
Self-discipline	0.144	**0.383**	0.121	0.012
Diligence	0.123	0.121	0.108	0.020

H. *Model Evaluations and Interpretations*

We estimated machine learning algorithms' predictive quality and reliability using established metrics such as F1-Score, precision, sensitivity, accuracy, confusion matrix, receiver operating characteristic (ROC), and area under the curve (AUC). The interpretation of information was accomplished by evaluating the influence of an attribute's inclusion or exclusion on a model via permutation feature importance, apriori algorithm for interesting patterns, and correlation analysis to assess the association between variables.

III. Results

The succeeding section presents the hyperparameter configurations, model evaluation and interpretations, comparative consistencies, and correlational analysis results.

A. *Hyperparameter Tuning*

In contrast to the learned parameters of a model, hyperparameters are user-specified. While providing suitable values improves the performance of machine learning, it is one of the most time-consuming and underutilized tasks in knowledge discovery. Although it is a computationally intensive technique, the benefit outweighs the drawbacks.

TABLE 4
CLASSIFICATION MODEL'S HYPERPARAMETER SETTINGS

Model	Configurations
J48	splitter = 'best', min_samples_leaf = 1, max_leaf_nodes = 5, max_depth = 4
LogReg	penalty = 'l2', solver = 'newton-cg'
ANN	activation = 'relu', solver = 'adam', batch_size = 50
KNN	metric = 'manhattan', weights = 'uniform', n_neighbors = 8
RF	n_estimators = 25, max_features = 'log2', max_depth = 4
SVM	gamma = 0.01, kernel = 'rbf', C = 100
NB	var_smoothing = 0.0218

B. *Evaluation Scores*

Table 5 demonstrates the Random Forests' robustness with an accuracy of 92.09% because it uses multiple ensemble trees capable of solving over fitting [29], followed by Artificial Neural Network and J48 Decision Tree with 88.41% and 85.40%. Support Vector Machines, Naïve Bayes, and K-Nearest Neighbor performed moderately with 84.67%, 83.33%, and 81.88%, while the Logistic Regression underperforms with 81.16%.

TABLE 5
CROSS-VALIDATED SCORES (10-FOLD)

Model	Accuracy	Precision	Sensitivity	F1 - Score
RF	**0.9209**	0.9429	0.9041	0.9231
ANN	0.8841	0.8986	0.8732	0.8857
J48	0.8540	0.8676	0.8429	0.8551
SVM	0.8467	0.8676	0.8310	0.8489
NB	0.8333	0.8406	0.8286	0.8345
KNN	0.8188	0.8056	0.8406	0.8227
LogReg	0.8116	0.8406	0.7945	0.8169

C. *Confusion Matrix*

Table 6 exhibits the distribution of correct and incorrect classification of the RF. A confusion matrix is a visual representation and summary of the classifier's performance. Based on the results, it signifies that the prediction of gritty individuals (93.75%) is more accurate than non-gritty ones (88.70%).

TABLE 6
RANDOM FORESTS' CONFUSION MATRIX

Test Size, N = 139 (30%)	Predicted: High Grit	Predicted: Low Grit
Actual: High Grit	64	4
Actual: Low Grit	7	62

Furthermore, we measure the model's predictive quality and discriminative ability using ROC and AUC curve plots. The graph determines the capacity of the model to distinguish between classes – higher scores for both plots mean better performance. Figure 4 illustrates an AUC rating of 0.979 accounting for a non-random prediction beyond the guessing (red) line.

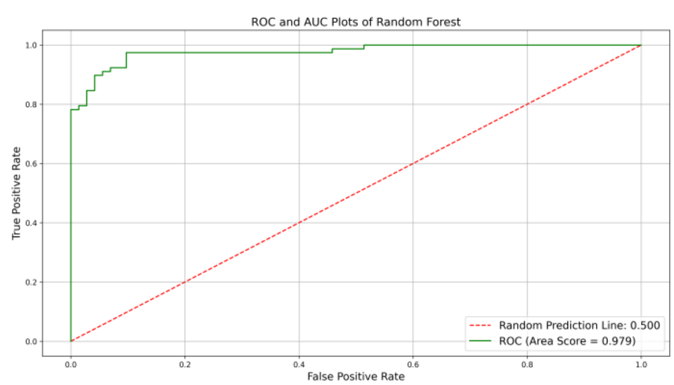

Fig. 4. ROC and AUC plots of the model's discriminative ability

D. *Model Interpretation*

'Black box' models such as Random Forests are complicated to decipher because the insight is buried within its probabilistic structure, making it impossible to draw out the relationships. Our solution is to use permutation feature importance. The procedure disrupts the link between class labels and predictors by randomly exchanging variable values.

A reduction of a model's average prediction accuracy indicates the attribute's significance. Table 7 shows the estimations using ten cross-validated arbitrary trials.

TABLE 7
PERMUTATION FEATURE IMPORTANCE

Weight	Feature
0.2587 ± 0.0813	Perseverance
0.1617 ± 0.0413	Sustained interests
0.1131 ± 0.0392	Self-motivation
0.0715 ± 0.0281	Self-discipline

The results identified that perseverance (0.2587) was the best predictor, followed by sustained interest (0.1617), self-motivation (0.1131), and self-discipline (0.0715). In addition, we implemented association rule mining (apriori) based on each attribute's average values (high & low baselines) for discovering interesting relationships, frequent associations, and causality between characteristics [30]. Table 8 shows the consequent and antecedents with a 0.5 support configuration.

TABLE 8
INTERESTING RELATIONSHIPS AND FREQUENT ASSOCIATIONS OF GRIT

Levels	Frequent Patterns and Associations	Support
High grit	(High perseverance)	**0.935**
	(High sustained interest)	0.862
	(High perseverance, High sustained interest)	0.843
	(High perseverance, High self-motivation)	0.617
	(High self-motivation, High self-discipline)	0.506
Low grit	(Low perseverance)	**0.949**
	(Low sustained interest)	0.885
	(Low self-motivation, Low sustained interest)	0.794
	(Low perseverance, Low self-motivation)	0.715
	(Low self-motivation, Low self-discipline)	0.625

It shows that grit levels are positively linked to perseverance with support of 0.935 (high grit) and 0.949 (low grit). The second indicator was the sustained interest (0.862 & 0.885). Students with low grit show combinations of low self-motivation and sustained interests (0.794), low perseverance and low self-motivation (0.715), and low self-motivation and low self-discipline (0.625)

E. *Consistency of Clustering and Classification*

Table 9 summarized the percentage consistency of the clustering against the classification for 500 learners. This test adds to a model's validation verification. We observed a higher score for the 'high grit' class with 97.63% while 97.61% for the 'low grit' groups. Both algorithms categorization are not far off, and the calculated consistencies are highly acceptable.

TABLE 9
CONSISTENCY RESULTS FOR DBSCAN AND RANDOM FORESTS

Group/Class Labels	Algorithms		% Consistency
	Clustering (DBSCAN)	Classification (Random Forests)	
High grit	248	254	97.63%
Low grit	252	246	97.61%

F. *Consistency of Grit Scale Questionnaire vs. Machine*

Table 10 summarized the percentage consistency of the grit scale questionnaire (>= 4.1 is high & < 4.1 is low) [31] against both clustering and classification for the entire 500 students. We found higher consistency scores for Random Forests with 94.07% and 93.49%. On the other hand, DBSCAN yields 91.85% and 90.55%. All are acceptable consistency ratings.

TABLE 10
CONSISTENCY RESULTS OF GRIT SCALE QUESTIONNAIRE VS. MACHINE

Group/ Class Labels	Algorithms		Survey	% Consistency	
	Clustering (DBSCAN)	Classification (Random Forests)	Grit Scale	DBSCAN vs. Grit Scale	RF vs. Grit Scale
High grit	248	254	270	91.85%	94.07%
Low grit	252	246	230	90.55%	93.49%

G. *Correlation of Grit Attributes to Course Ratings and Professional Achievement*

Correlation is a statistical method for examining the relationship strengths between two variables and calculating their associations. We opted to utilize a Spearman rank correlation instead of a Pearson's *r* due to the nonparametric characteristics of our data (e.g., not normally distributed, see Figure 2). Based on Table 11, there is a generally weak correlation [32] between the course rating with self-motivation (0.308), and self-discipline (0.302); however, perseverance (0.393) and sustained interest (0.345) are not significant at 0.05 levels.

TABLE 11
GRIT ATTRIBUTES VS. COURSE RATING CORRELATION (N = 500)

Attributes	Course Rating [a] (correlation (p-value))
Perseverance	0.393 (0.057)
Sustained interest	0.345 (0.058)
Self-motivation	0.308 (0.048)
Self-discipline	0.302 (0.043)

a. Outstanding (1.00–1.24), Very Satisfactory (1.25-1.74), Satisfactory (1.75-2.49), Fair (2.50–2.99), Pass (3.00)

In stark contrast, professional achievement (based on earnings) dictates a very strong correlation [32] with perseverance (0.914) and sustained interests (0.908), but self-motivation (0.735) and self-discipline (0.713) are not significant at 0.05 (see Table 12). Tracer study reported that 273 of 500 (54.60%) graduates with high-level combinations of perseverance and sustained interests based on their e-learning behaviors have a high-paying profession above the country's average wages. Only 29 (5.80%) graduates have attained high salaries with the opposite characteristics.

TABLE 12
GRIT ATTRIBUTES VS. PROFESSIONAL ACHIEVEMENT CORRELATION (N = 500)

Attributes	Professional Achievement [a] (correlation (p-value))
Perseverance	0.914 (0.0008)
Sustained interest	0.908 (0.015)
Self-motivation	0.735 (0.058)
Self-discipline	0.713 (0.056)

a. Professional achievement according to job levels: Senior (3), Intermediate (2), and Associate (1) [33]

IV. DISCUSSIONS

This study exemplified the capacity of machine learning to identify grit in an e-learning environment using multiple modeling techniques. Our results clearly illustrated that intuitive data mining strategies such as dynamic time warping, clustering, feature scaling, dimensionality reduction, hyperparameter optimization, and classification augment prediction accuracy. The study also noted that it is best to complement rather than compare the strengths of unsupervised and supervised learning. Permutation feature importance and association rule mining analysis ranked perseverance and sustained interest as the two most salient components of grit in agreement with [34 - 36]. Furthermore, generalization robustness verification of our model yields an acceptable consistency score of 92.67% against the manual method, confirming the validity of the grit scale questionnaire. This work disclosed a weak correlation between course rating and grit attributes, confirming the articles of [37] [38], whereas professional achievement shows a strong positive correlation [4] [39]. Moreover, it exhibits that, while grit is firmly related to long-term goals, results revealed that it might not be the case for short-term academic achievement such as course grades. The lack of correlation between the two variables is attributed to age-related changes in an individual's grit levels, similar to the findings by [13]. It suggests that grit may lack discriminant validity in younger people because they still discover their interests and continually change. Their persistent efforts are not yet well-observed [37].

From a developmental perspective, individuals experience an identity crisis during adolescence [40]. They interact with other people and learn new roles and ideas that may influence their self-concept. As a result, they question who they should be, what they are good at, and what they want to be in the future, which explains their interests' dynamic state. The author [41] also explained that when young adults enter early adulthood, part of their tasks is to establish their career, start rearing a family, and have more commitment to long-term goals [42]. Their interests are more stable at this stage, and their persistent efforts in their careers are evident. Thus, this research attested that developmental changes and maturational variables could be considered significant in explaining why grit attributes are linked with long-term than short-term goals. Our study is limited only to discernible variables observed in the confines of e-learning and did not include other essential characteristics such as demographics.

V. Conclusions and Future work

Not all accomplishments are purely the consequence of intellectual aptitude and talent. Instead, intangibles such as perseverance and resilience are critical factors deciding one's success. The rise of big data in academic settings gives rise to new data-driven approaches to enhance initiatives in educational efficacy. Learning and behavioral processes can now be studied in detail using students' digital activity traces, which previously were too expensive to obtain via conventional sources. Our research discovered the potential of determining learners' grit through multiple stages of machine learning approaches and extensive modeling. More importantly, we contributed and shed light on the relationship of grit with academic and professional achievement. Although grit is not strongly associated with short-term academic success, it is adamantly connected with professional success due to maturation. This study's empirical evidence demonstrates the critical nature of cultivating non-cognitive abilities such as grit among students, which has the predictive ability for long-term achievements. Our investigation has profound implications for various educational institutions that holistic development of students' (cognitive & non-cognitive) skills must be given full attention. The authors plan to improve the model's accuracy in the future by experimenting with alternative algorithms and including significant additional data.

Acknowledgement

The lead author expresses his gratitude for the contributions of the co-authors, the Commission on Higher Education (CHED), and Southern Luzon State University (SLSU) for supporting and funding this research through institutional development and innovation grants (IDIG).

References

[1] D. Park, A. Yu, R. N. Baelen, E. Tsukayama and A. L. Duckworth, "Fostering grit: Perceived school goal-structure predicts growth in grit and grades", Contemporary Educational Psychology, vol. 55, pp. 120-128, 2018.

[2] M. Christopoulou, A. Lakioti, C. Pezirkianidis, E. Karakasidou and A. Stalikas, "The role of grit in education: A systematic review", Psychology, 9(15), pp. 2951-2971, 2018.

[3] J. M. Stoffel and J. Cain, "Review of grit and resilience literature within health professions education", American Journal of Pharmaceutical Education, 82(2), pp. 124-134, 2018.

[4] A. L. Duckworth, C. Peterson, M. D. Matthews and D. R. Kelly, "Grit: perseverance and passion for long-term goals", Journal of Personality and Social Psychology, 92(6), pp. 1087-1101, 2007.

[5] J. M. Jachimowicz, A. Wihler, E. R. Bailey and A. D. Galinsky, "Why grit requires perseverance and passion to positively predict performance", Proceedings of the National Academy of Sciences, 115(40), pp. 9980-9985, 2018.

[6] P. Akos and J. Kretchmar, "Investigating grit at a non-cognitive predictor of college success", The Review of Higher Education, 40(2), pp. 163-186, 2018.

[7] R. S. Clark and V. L. Plano Clark, "Grit within the context of career success: A mixed methods study", International Journal of Applied Positive Psychology, 4(3), pp. 91-111, 2019.

[8] R. R. Maaliw, K. A. C. Quing, A. C. Lagman, B. H. Ugalde, M. A. Ballera and M. A. D. Ligayo, "Employability prediction of engineering graduates using ensemble classification modeling", IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), pp. 288-294, 2022.

[9] M. Aparicio, F. Bacao and T. Oliveira, "Grit in the path to e-learning success", Computers in Human Behavior, vol. 66, pp. 388-399, 2017.

[10] R. E. Allen, C. Kannangara and J. Carson, "True grit: How important is the concept of grit for education? A narrative literature review", International Journal of Educational Psychology: IJEP, 10(1), pp. 73-87, 2021.

[11] A. L. Duckworth and P. D. Quinn, "Development and validation of the short grit scale (GRIT–S)", Journal of Personality Assessment, 91(2), pp. 166-174, 2009.

[12] K. Fosnacht, K. Copridge and S. A. Sarraf, "How valid is grit in the postsecondary context? A construct and concurrent validity analysis", Research in Higher Education, 60(6), pp. 803-822, 2019.

[13] C. S. Kannangara, R. E. Allen, G. Waugh, N. Nahar, S. Z. N. Khan, S. Rogerson and J. Carson, "All that glitters is not grit: Three studies of grit

in university students”, Frontiers in Psychology, 9(1539), pp. 1-15, 2018.

[14] M. Credé, “Total grit scale score does not represent perseverance”, Proceedings of the National Academy of Sciences, 116(10), pp. 3941-3941, 2019.

[15] C. J. Fong and Y. W. Kim, “A clash of constructs? Re-examining grit in light of academic buoyancy and future time perspective”, Current Psychology, 40(4), pp. 1824-1837, 2021.

[16] O. Gonzalez, J. R. Canning, H. Smyth and D. P. MacKinnon, “A psychometric evaluation of the short grit scale”, European Journal of Psychological Assessment, 2019.

[17] R. De Koning, A. Egiz, J. Kotecha, A. C. Ciuculete, S. Z. Y. Ooi et al., “Survey fatigue during the COVID-19 pandemic: an analysis of neurosurgery survey response rates”, Frontiers in Surgery, 8(690680), pp. 1-15, 2021.

[18] R. R. Maaliw and M. Ballera, “Classification of learning styles in virtual learning environment using J48 decision tree”, Proceedings of the 14th International Conference on Cognition and Exploratory Learning in Digital Age, Algarve, Portugal, pp. 146-156, 2017.

[19] R. R. Maaliw, “A personalized virtual learning environment using multiple modeling techniques”, IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), pp. 8-15, 2021.

[20] X. Tang, M. T. Wang, J. Guo and K. Salmela-Aro, “Building grit: The longitudinal pathways between mindset, commitment, grit, and academic outcomes”, Journal of Youth and Adolescence, 48(5), pp. 850-863, 2019.

[21] É. P. Caillault, A. Lefebvre and A. Bigand, “Dynamic time warping-based imputation for univariate time series data”, Pattern Recognition Letters, 139, pp. 139-147, 2020.

[22] A. Lahreche and B. Boucheham, “A fast and accurate similarity measure for long time series classification based on local extrema and dynamic time warping”, Expert Systems with Applications, 168, 114374, 2021.

[23] E. Schubert, J. Sander, M. Ester, H. Kriegel and X. Xu, “DBSCAN revisited: Why and how you should (still) use DBSCAN”, ACM Transactions on Database Systems (TODS), 42(3), pp. 1-21, 2017.

[24] D. Luchi, A. L. Rodrigues and F. M. Varejão, “Sampling approaches for applying DBSCAN to large datasets”, Pattern Recognition Letters, 117, pp. 90-96, 2019.

[25] K. R. Shahapure and C. Nicholas, “Cluster quality analysis using silhouette score”, IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), pp. 747-748, 2020.

[26] F. Chamroukhi and H. D. Nguyen, “Model‐based clustering and classification of functional data”, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(4), e1298, 2019.

[27] B. H. Shekar and G. Dagnew, “Grid search-based hyperparameter tuning and classification of microarray cancer data”, 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), pp. 1-8, 2019.

[28] S. Karamizadeh, S. M. Abdullah, A. A. Manaf, M. Zamani and A. Hooman, “An overview of principal component analysis”, Journal of Signal and Information Processing, 2020.

[29] M. Belkin, D. J. Hsu and P. Mitra, “Overfitting or perfect fitting? Risk bounds for classification and regression rules that interpolate”, Advances in Neural Information Processing Systems, vol. 31, 2018.

[30] A. Telikani, A. Gandomi and A. Shahbahrami, “A survey of evolutionary computation for association rule mining”, Information Sciences, vol. 524, pp. 318-352, 2020.

[31] M. Cazayoux, A. Bishop, J. Navalta, C. Harris, K. Adams and M. DeBeliso, “The reliability of the 12-item grit scale among crossfit participants”, European Journal of Physical Education and Sport Science, 2018.

[32] P. Schober, C. Boer and L. A. Schwarte, “Correlation coefficients: Appropriate use and interpretation”, Anesthesia & Analgesia, 126(5), pp. 1763-1768, 2018.

[33] Salaries, allowances, benefits, and job classifications in the Philippines. Accessed: January 29, 2022, [Online], Available: https://www.un.org/Depts/OHRM/salaries_allowances/salaries/philippi.htm

[34] J. M. Jachimowicz, A. Wihler, E. R. Bailey and A. D. Galinsky, “Why grit requires perseverance and passion to positively predict performance”, Proceedings of the National Academy of Sciences, 115(40), pp. 9980-9985, 2018.

[35] E. Sudina and L. Plonsky, “Academic perseverance in foreign language learning: An investigation of language-specific grit and its conceptual correlates”, The Modern Language Journal, 2021.

[36] M. S. Hagger and K. Hamilton, “Grit and self-discipline as predictors of effort and academic attainment”, British Journal of Educational Psychology, 89(2), pp. 324-342, 2019.

[37] P. Bazelais, D. J. Lemay and T. Doleck, “How does grit impact college students’ academic achievement in science”, European Journal of Science and Mathematics, 4(1), pp. 33-43, 2017.

[38] A. J. L. Palisoc, R. R. Matsumoto, J. Ho, P. J. Perry, T. T. Tang and E. J. Ip, “Relationship between grit with academic performance and attainment of postgraduate training in pharmacy students”, American Journal of Pharmaceutical Education, 81(4), pp. 1-10, 2017.

[39] N. Brooks and S. J. Seipel, “Grit and the information systems student: A discipline-specific examination of perseverance and passion for long term goals”, Information Systems Education Journal, 16(1), pp. 21-32, 2018.

[40] M. Block, “Identity versus role confusion”, Encyclopedia of Child Behavior and Development, Springer, Boston, MA., 2011.

[41] E. R. Freitas, A. J. G. Barbosa, N. Scoralick-Lempke, N. Magalhães et al., “Developmental tasks and life history of the elderly: Analysis of Havighurst’s theory”, Psicologia, Reflexão e Crítica, 26(4), pp. 809-819, 2013.

[42] K. Hurrelman and G. Quenzel, “Lost in transition: Status insecurity and inconsistency as hallmarks of modern adolescence”, International Journal of Adolescence and Youth, 20(3), pp. 261-270, 2015.

2022 IEEE World AI IoT Congress, AIIoT 2022 • Pages 39 - 45 • 2022 • 2022 IEEE World AI IoT Congress, AIIoT 2022 • Seattle • 6 June 2022through 9 June 2022 • Code 180897

Document type
Conference Paper
Source type
Conference Proceedings
ISBN
978-166548453-4
DOI
10.1109/AIIoT54504.2022.9817177
View more

Clustering and Classification Models for Student's Grit Detection in E-Learning

Maaliw, Renato R.[a] ; Quing, Karen Anne C.[b] ; Susa, Julie Ann B.[a] ; Marqueses, Jed Frank S.[b] ; Lagman, Ace C.[c] ; Adao, Rossana T.[d] ; Fernando-Raguro, Ma. Corazon[c] ; Canlas, Ranie B.[e]

Save all to author list

[a] College of Engineering Southern Luzon State University, Lucban, Philippines
[b] College of Arts and Sciences Southern Luzon State University, Lucban, Philippines
[c] FEU Institute of Technology, Information Technology Dept., Manila, Philippines
[d] College of Computer Studies FEU Institute of Technology, Manila, Philippines
View additional affiliations

Abstract

Grit plays a crucial role in determining high individual success more than intellectual talent alone. However, there is no existing literature that ventured into the trait identification in an e-learning environment. This study presents a comprehensive computational-driven strategy for detecting a learner's grit using machine learning. Empirical results show that DBSCAN and Random Forest models produce average accurate prediction consistency of 92.67% against the questionnaire method. Knowledge interpretation using feature importance and association mining quantifies perseverance and sustained interest as the most pressing component of grit. Correlational analysis reveals that grit has a weak connection with course grades (short-Term goal) but demonstrates a strong positive association with professional achievement (long-Term goal) and maturation. Collectively, our findings substantiate that breakthrough accomplishment is contingent not solely on cognitive ability but on constant interests and resilience. copy; 2022 IEEE. © 2022 IEEE.

Author keywords

classification; density-based clustering; dynamic time warping; grit scale; non-cognitive traits; prediction

Indexed keywords

Sustainable Development Goals 2023 New

SciVal Topics

Metrics

Funding details

Sustainable Development Goals 2023
SciVal Topics
Metrics
Funding details

References (42)

1 Park, D., Yu, A., Baelen, R.N., Tsukayama, E., Duckworth, A.L.

Clustering and Classification Models For Student's Grit Detection in E-Learning

Publisher: IEEE Cite This PDF

Renato R. Maaliw ; Karen Anne C. Quing ; Julie Ann B. Susa ; Jed Frank S. Marqueses ; A... All Authors

Abstract:
Grit plays a crucial role in determining high individual success more than intellectual talent alone. However, there is no existing literature that ventured into the trait identification in an e-learning environment. This study presents a comprehensive computational-driven strategy for detecting a learner's grit using machine learning. Empirical results show that DBSCAN and Random Forest models produce average accurate prediction consistency of 92.67% against the questionnaire method. Knowledge interpretation using feature importance and association mining quantifies perseverance and sustained interest as the most pressing component of grit. Correlational analysis reveals that grit has a weak connection with course grades (short-term goal) but demonstrates a strong positive association with professional achievement (long-term goal) and maturation. Collectively, our findings substantiate that breakthrough accomplishment is contingent not solely on cognitive ability but on constant interests and resilience.

Published in: 2022 IEEE World AI IoT Congress (AIIoT)

Date of Conference: 06-09 June 2022

Date Added to IEEE *Xplore*: 13 July 2022

ISBN Information:
Electronic ISBN:978-1-6654-8453-4
Print on Demand(PoD) ISBN:978-1-6654-8454-1

INSPEC Accession Number: 21864664

DOI: 10.1109/AIIoT54504.2022.9817177

Publisher: IEEE

Conference Location: Seattle, WA, USA

Funding Agency:

I. Introduction

The significance of intellectual capacity is well documented in academic and professional domains, but we understood very little about other equally critical characteristics that foresee success. While various demographic features have a role in both contexts, many including personality and behavioral factors. Stu gnitive ability is insufficient [1]. They must surmount hurdles, be dedicated and disciplined, and maintain high motivation to achieve goals while remaining resilient in any

Authors

Renato R. Maaliw
College of Engineering Southern Luzon State University, Lucban, Quezon, Philippines

Karen Anne C. Quing
College of Arts & Sciences Southern Luzon State University, Lucban, Quezon, Philippines

Julie Ann B. Susa
College of Engineering Southern Luzon State University, Lucban, Quezon, Philippines

Jed Frank S. Marqueses
College of Arts & Sciences Southern Luzon State University, Lucban, Quezon, Philippines

Ace C. Lagman
Information Technology Dept., FEU Institute of Technology, Manila, Philippines

Rossana T. Adao
College of Computer Studies FEU Institute of Technology, Manila, Philippines

Ma. Corazon Fernando - Raguro
Information Technology Dept., FEU Institute of Technology, Manila, Philippines

Ranie B. Canlas
College of Computer Studies Don Honorio Ventura State University, Bacolor, Pampanga, Philippines

Figures

References

Citations

Keywords

Metrics

My... »My profile »#69 (1570805872): Clustering and Classification Models for Student's Grit Detection in E-Learning

#69 (1570805872): *Clustering and Classification Models for Student's Grit Detection in E-Learning*

BibTeX

Property	Change Add	Value
Conference and *track*		**2022 IEEE World AI IoT Congress (AIIoT)** - *2022 IEEE World AI IoT Congress (AIIoT) Regular Research Paper*

Authors

Name	ID	Edit	Flag	Affiliation (edit for paper)	Email	Country	Register
Renato R. Maaliw III	1841045			Southern Luzon State University, Philippines	renatomaaliw3@gmail.com	Philippines	
Karen Anne C. Quing	1950145			Southern Luzon State University, Philippines	kquing4569@gmail.com	Philippines	
Julie Ann B. Susa	1950146			Southern Luzon State University, Philippines	jsusa4569@yahoo.com	Philippines	
Jed Frank S. Marqueses	1950148			Southern Luzon State University, Philippines	jmarqueses4569@yandex.com	Philippines	
Ace C. Lagman	1925232			FEU Institute of Technology, Philippines	acelagman69@gmail.com	Philippines	
Rossana T. Adao	1950150			FEU Institute of Technology, Philippines	radao4569@yandex.com	Philippines	
Ma. Corazon Fernando - Raguro	1943708			FEU Institute of Technology, Philippines	mcorazon4569@gmail.com	Philippines	
Ranie B. Canlas	1950152			Don Honorio Ventura State University, Philippines	rcanlas4569@yandex.com	Philippines	

Property	Change Add	Value
Title	Only the chairs (ieeeaiiot2022-chairs@edas.info) can edit	*Clustering and Classification Models for Student's Grit Detection in E-Learning*
Abstract	Only the chairs (ieeeaiiot2022-	Grit plays a crucial role in determining high individual success more than intellectual talent alone. However, there is no existing literature that ventured into the trait identification in an e-learning environment. This study presents a comprehensive computational-driven strategy for detecting a learner's grit using machine learning. Empirical results show that our clustering and classification models produce average accurate prediction consistency of 90.43%. Knowledge interpretation using feature importance and association mining

	chairs@edas.info) can edit	quantifies perseverance and sustained interest as the most pressing component of grit. Correlational analysis reveals that grit has a weak connection with course grades (short-term goal) but demonstrates a strong positive association with professional achievement (long-term goal) and maturation. Collectively, our findings substantiate that breakthrough accomplishment is contingent not solely on cognitive ability but on constant interests and resilience.
Keywords	Only the chairs (ieeeaiiot2022-chairs@edas.info) can edit	classification; density-based clustering; dynamic time warping; grit scale; non-cognitive traits; prediction
Topics	Only the chairs (ieeeaiiot2022-chairs@edas.info) can edit	Big Data, Data Management and Analytics
Presenter(s)		Karen Anne C. Quing, Jed Frank S. Marqueses and Renato R. Maaliw III (bio)
Registration		Renato R. Maaliw III has registered (but not paid) for T2:IEEEM
DOI	Only the chairs (ieeeaiiot2022-chairs@edas.info) can edit	
URL		
Status		Accepted
Copyright form		IEEE; IEEE: May 20, 2022 07:04 Asia/Hong_Kong
Review manuscript		However, authors cannot upload: paper status

Document (show)	Pages	File size	Changed
	7	480,692	Apr 26, 2022 18:50 Asia/Hong_Kong

Can upload 10 pages (type) until May 27, 2022 07:59 Asia/Hong_Kong.

Final manuscript

Document (show)	Pages	File size	Changed	Check format / Report problem	Delete
	7	467,580	May 22, 2022 18:59 Asia/Hong_Kong	checked May 22, 2022 18:59 Asia/Hong_Kong	

pdf	abstractsdiffer	The abstract in the manuscript file differs from the paper meta data (95.93). (FAQ 1084)	-

Presentation	Can upload any length (type) until Jun 2, 2022 07:59 Asia/Hong_Kong.
Stamped for attendee proceedings	However, authors cannot upload:

Personal notes

You are the creator, an author and a reviewer for this paper. You have authored an accepted paper.

Reviews

3 IEEE AIIoT 2022s

Review 1

None of the scores are visible for your roles.

Comments (Please write 100 words regarding this review report)

Interesting idea, but not innovative. Similar work:
https://www.tandfonline.com/doi/abs/10.1080/10494820.2021.1983611
I can see this idea being expanded for other academic and similar professions as well.

Only one little question, in section A of methodology, the authors selected 500 most active account, and later on it says 300 randomly selected learners?

Review 2

None of the scores are visible for your roles.

Comments (Please write 100 words regarding this review report)

The authors present an interesting study that uses data from an on-line course to classify the level of grit of students.
They use several different machine learning techniques.
The results are interesting and they may be the basis for additional
research and reflection on the part of educators.
There are important questions that will need to be addressed by future research: Are there certain environments that are more attractive and engaging for certain types of students than others? Perhaps what is being evaluated is not necessarily the grit, as a personal trait, but the match between a learning environment and the expectations of a student. Perhaps it is a combination of both?
Could this information be used to advice students about their academic choices?

Review 3

None of the scores are visible for your roles.

Comments (Please write 100 words regarding this review report)

The research is written in a professional and organized manner,
But the acknowledgment part is blank. That should be filled.
A careful review of the previous research analysis is required.
The graphs should be in a good resolution.

FACULTY POSITION RECLASSIFICATION FOR SUCS
(DBM-CHED Joint Circular No. 3, series of 2022)

CERTIFICATION OF PERCENTAGE CONTRIBUTION

(Research Output with Multiple Authors)

Title of Research: Clustering and Classification Models For Student's Grit Detection in E-Learning
Type of Research Output: Conference Paper (Scopus-Indexed, IEEE Paper)

Instruction: Supply ALL the names of the authors involved in the publication of Research output and indicate the contribution of each author in percentage. Each author shall sign the conforme column if he/she agrees with the distribution. The conforme should be signed by all the authors in order to be considered. Please prepare separate Certification for each output.

	Name of Authors	% Contribution	Conforme *(Sign if you agree with the % distribution)*
1	Renato R. Maaliw III	25%	
2	Karen Anne C. Quing	40%	
3	Julie Ann B. Susa	10%	
4	Jed Frank S. Marqueses	5%	
5	Ace C. Lagman	5%	
6	Rossana T. Adao	5%	
7	Ma. Corazon F. Raguro	5%	
8	Ranie B. Canlas	5%	
	** Should have a total of 100%*	100.00%	

Prepared by:

Renato R. Maaliw III, DIT
(Name and Signature)
Faculty

Certified by:

Nicanor L. Guinto, Ph.D
(Name and Signature)
Director, Office of Research Services