INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

Uncovering the Most Effective Pedagogical Techniques for Math Education Using Machine Learning

Said Elnaffar [a,*], Mohamed Fawey [b]

[a] Computer Science Department, Canadian University Dubai, Al Safa Street, Dubai, 117781, United Arab Emirates
[b] College of Computing and I, Arab Academy for Science, Technology and Maritime Transport, Egypt
Corresponding author: *said.elnaffar@cud.ac.ae

Abstract—**Many new math educators express that their first years in the teaching field are extremely challenging. They struggle to discover and apply the most effective teaching techniques and behaviors, often without the support of more experienced colleagues. In this study, we use machine learning to find the strategies that novice teachers can adopt to enhance their teaching effectiveness. The core aim is to uncover the relationship between teachers' performance, as assessed by student evaluations, and their pedagogical methods. These strategies are derived from the final decision tree model, which is trained on a large dataset of empirical data from schools. The data consists of input from 72 math teachers of grades 7-9 and their students in Dubai, used to train two decision tree models: a classification tree and a regression tree. The structure of these trees is analyzed to identify and rank the effectiveness of nine teaching techniques—such as Visualization, Practice, Math Rules, Gamification, Collaboration, Problem-Solving, Case Studies, Assessments, and Language Switching—and four behavioral methods—such as Inspiration, Engagement, Entertainment, and Bonding— in relation to the Student Evaluation Index (SEI), which is derived from student feedback. Results indicate that techniques such as "gamification" and "inspirational behavior" are consistently associated with higher SEI scores across different tree configurations. However, factors such as the demographics and culture of both students and teachers may need to be considered when generalizing these findings to other regions of the world.**

Keywords—**Education; machine learning; decision tree; mathematics; professional development.**

I. INTRODUCTION

The UNESCO Institute of Statistics stated in 2016 that the world must recruit 68.8 million new teachers to provide every future child with primary and secondary education [1]. Current strategies to assist new teachers require many man-hours from both new and experienced educators. Understanding the best teaching techniques is considered critical in intermediate education. This is especially true for fundamental subjects such as mathematics, as it paves the way for later subjects. Research shows that teaching mathematics is a challenging task [2]. This can be attributed to many teachers not knowing or using suitable techniques to teach mathematics at a high level. Additionally, Fadlelmula et al. [3] states that there exist social and psychological barriers including fear of failure, lack of confidence, and lack of supportive colleagues.

These obstacles may affect teachers' performance, leading to students not mastering essential mathematics concepts.

This could prevent students from pursuing STEM (science, technology, engineering, and mathematics) careers later in life that not only benefit society, but their own potential livelihood as well.

Support networks and continuing education programs for educators are inconsistently effective in helping teachers develop further. Hammond et al. notes some deficiencies of current professional development (PD) for teachers and analyzes some common features of successful PD programs [4]. The research conducted in our paper seeks to explore a specific direction of PD which is optimizing the content which includes pedagogical techniques and behaviors.

Researchers and practitioners from different fields seek solutions via machine learning techniques including sentiment analysis [24], [27], security and phishing attack detection [25], [26], health-related issues such as calories burn prediction [28] and ECG heartbeat classification [30], prediction of travel insurance purchases [29], and more.

Applications of machine learning (ML) have proliferated to many fields including education [5]-[7]. Some common applications include decoding students' problems, task automation and performance evaluation [8]. Machine learning carries the benefit of being able to capture and distill trends in vast quantities of data. As such, they may be able to improve or augment traditional professional development programs due to the vast amount of information they can generalize and extrapolate. Some researchers tried to predict the dropout for college students in MOOCs based on weighted multi-features [23], while others built proposed solutions for class imbalance as they tried to predict students graduating in time [31]. A machine learning model can reduce the hours required to provide new teachers with invaluable information to streamline their PD. This work seeks to convert real life teacher experience into data to be represented by decision tree models which are utilized to assist future teachers in the classroom.

The main contribution of our work is the identification of the most effective techniques that new teachers can use to enhance their teaching performance. These techniques are recommended by the final decision tree model trained on a substantial volume of empirical data from schools. In the following subsections we give a background about the traditional and contemporary teaching techniques, supervised machine learning, decision trees, and the principal Component Analysis (PCA) method.

A. Traditional Teaching Techniques

Classical teaching methods have their share of advantages [8] such as:

1) Cost-effectiveness: Conventional instructional methods, including in-person teaching, are more affordable compared to contemporary educational techniques. This affordability allows them to be adopted by rural areas without the burden of excessive costs.

2) Efficiency in Learning: The focus of conventional educational settings is on the swift conveyance of a large volume of information within a minimal timeframe.

3) Efficiency in Learning: The focus of conventional educational settings is on the swift conveyance of a large volume of information within a minimal timeframe.

4) Suitability for Certain Subjects: Subjects like physics, chemistry, and mathematics often benefit from direct instruction using a blackboard, as concepts in these fields can be more easily grasped through visual chalkboard demonstrations.

5) Suitability for Certain Subjects: Subjects like physics, chemistry, and mathematics often benefit from direct instruction using a blackboard, as concepts in these fields can be more easily grasped through visual chalkboard demonstrations.

6) No Need for Advanced Technology: Conventional teaching methods do not require educators to be skilled in modern technological tools, such as incorporating computers into classroom instruction.

Nevertheless, the advantages of traditional teaching methods are counterbalanced by their disadvantages. Teachers frequently focus on extensive note-taking and rote memorization, which can hinder a deep comprehension of fundamental concepts. This approach, aimed at delivering maximum information, often leads to a loss of student interest and understanding. Additionally, traditional classrooms often lack real-life application examples, resulting in students feeling disengaged from the material being taught [9].

B. Contemporary Teaching Techniques

Many investigators found in novel technological ideas a means to support education. For example, during challenging times, such as the COVID-19 pandemic, some researchers [32] turned to technology to help young students learn Discrete Mathematics by building a platform called *LearnwithEmaa*. Over the course of a month, students interacted with an agent named *Emma* and reported significant improvements in their mathematical understanding and exam performance.

The modern or contemporary teaching approach emphasizes activities and places the learner at the center of curriculum planning and instruction. In this constructivist method, learners actively engage in the process to expand their knowledge and hone their skills [10]. Meanwhile, the mentor or instructor guides and supports students in focusing on the subject through interactive activities. This modern pedagogy fosters cooperation, reduces competition among students, and creates a healthy learning environment [11].

In recent years, the breadth of knowledge in science and technology has significantly increased, along with humanity's ability to adapt to new information in these fields. Consequently, there is a substantial demand for imaginative and creative minds to explore uncharted and unexplored areas across various professions. Education in the 21st century should be designed to prepare students for this era, which is driven by technological advancement and the development of individuals, societies, and nations. Modern teaching methods should be employed to educate future generations, providing them with the necessary information to fully capitalize on emerging opportunities [12], [13].

The coupling between teaching and learning, and artificial intelligence methods increased remarkably in the research community the recent years [22]. For example, the authors in [18] discussed the role of using digital technologies for teaching mathematics. Similarly, Wardat et al. [19] discussed the challenges, practices, and new perspectives of using AI in mathematics education. A systematic literature review of the use of technology in the pedagogy of mathematics was conducted by Aliyu et al. [20] [21].

C. Supervised Machine Learning

Machine learning (ML) involves developing algorithms that indirectly solve problems by detecting patterns in data. [14], [15]. ML algorithms or models can solve complex problems without explicit programming instruction that would directly affect the conclusions or solutions. Several notable use cases are in object detection, natural language processing, data forecasting and robotic path planning among others.
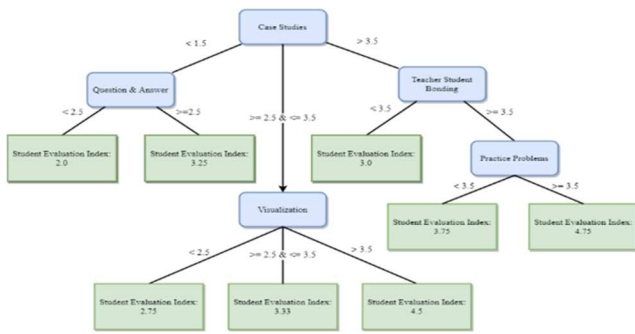
Fig. 1 A regression decision tree that predicts student performance given pedagogical techniques adopted.

ML attempts to perform actions based on previous experience which is represented by large datasets. Large amounts of data samples are required for training, which usually include input features and output labels. This is called Supervised Learning. However, there exist Unsupervised ML methods that only require inputs and will draw conclusions without knowledge of predefined answers. Supervised Learning is a subset of ML techniques that uses examples with input features and expected outputs values. For example, consider the simple function Equation 1.

$$Y=f(X) \tag{1}$$

The function f used to bind input features (X) to the desired outcome (Y) represents a machine learning model. Supervised methods will compare the predicted output of $f()$ to the expected output to generate an error value. This error value, depending on the ML algorithm, is used to update the parameters of the function to produce the desired behavior.

D. Decision Trees as a Classification Method

The decision tree method is a recursive algorithm for splitting data, applicable to both regression and classification. Despite its age and simplicity compared to techniques like neural networks, it's still a potent machine learning strategy. Asanbe et al. [16]. proves that decision trees used in the similar context of teacher performance evaluation from demographic, and education experience data can perform well.

A decision tree is a visual representation of decision-making scenarios that, when repeated, result in complex branching patterns (Fig. 1). These trees are utilized to derive insights from large datasets by establishing decision rules. The core principle involves dividing datasets using the most suitable decision rule for that set. These similar subsets are then recursively divided until all subsets are internally consistent or meet another criterion. This process creates a tree-like structure that can be navigated for predictive tasks. Decision trees are widely available in various data science programs, platforms, and machine learning libraries, making them easy to implement swiftly.

E. Principal Component Analysis (PCA) as a Dimensionality Reduction Method

Principal Component Analysis (PCA) is a method used to reduce the dimensionality of a dataset by identifying and eliminating correlations among features. It operates by computing the covariance of all input features to understand their interrelationships. Features that exhibit high correlations may be considered redundant and thus suitable for removal.

PCA transforms the data into new variables known as principal axes, which are derived from the dataset's eigenvectors. The principal axis that exhibits the most variance is deemed to contain the most information. By selecting the axes with the highest variance and discarding less informative ones, PCA simplifies the dataset.

This leads to the elimination of certain input features that contribute minimal new information. Implementing PCA can enhance model accuracy and reduce the time required for model training and inference. However, it also makes the task of extracting feature importance more challenging, as the original features are transformed and combined into new principal axes.

II. MATERIALS AND METHOD

A. Research Goals

In this research we aim to find answers to the following questions:
- Is it possible to predict public-school educator's effectiveness based on student satisfaction and academic achievement?
- Can teaching techniques or behaviors be ranked to guide educators in Dubai on which to adopt for maximizing student satisfaction?

Related work in teaching evaluation and prediction models applied to student performance indicate that machine learning evaluation of teaching methods can potentially be informative [13], [16].

B. Data Collection

A teacher self-assessment survey and a matching student evaluation of teachers were developed and voluntarily distributed to government school educators in Dubai to gather data.

The teacher survey gathers some demographic information about teachers such as location, school, relevant years of experience and so on as well as information on 13 techniques and behavioral methods described later in this section. The students' teacher evaluation survey consists of demographic information and the 3 student evaluation questions. To ensure the privacy of teachers and students, the survey results must not impact their careers or learning experiences. Neither survey gathered names or other identifying details. Teachers were asked to create a unique 4-character ID using letters from the English alphabet and numbers 0-9, as shown in the following example:

A89L

Codes such as this were distributed by teachers to their respective students. Teachers and students input this code into their surveys, which allowed the dataset to associate teacher responses to their respective classes while protecting teacher and student anonymity.

C. Target Population

The target populations are grades 7-9 mathematics teachers at government schools in Dubai. The following teachers were excluded from the study:
- Teachers who were not currently teaching in these grades but had previously done so.

- Teachers who were not actively teaching in the classroom.
- Part-time or student teachers.
- Teachers who did not teach mathematics full time.
- Private school teachers.

There were no restrictions on which students could take the student survey as long as they were a member of their mathematics teacher's class and were enrolled in grades 7-9.

D. Data Composition

The decision tree's features include scores for 4 behavioral methods and 9 teaching techniques, scored on a Likert Scale from 1 (never) to 5 (almost always). These behaviors and techniques, slightly modified from research on widely used mathematics pedagogies, cover both traditional and modern approaches [17]. Each teacher self-reported these techniques in their survey. The teaching techniques are as follows:

a. Visualization involves using visual tools to aid teaching.
b. Practice entails giving students homework and practical exercises.
c. Mathematical Rules means teaching through theoretical principles, rules, proofs, and laws.
d. Gamification uses games to communicate ideas and teach subjects.
e. Collaboration encourages students to work together in pairs or groups.
f. Solving Problems/Discovery involves imparting knowledge through storytelling or hands-on activities.
g. Case Studies/Scenarios present students with real-life or business-related applications or concepts.
h. Assessments evaluate students' knowledge through quizzes and feedback.
i. Language Switching involves using students' native language to ensure clear and accurate communication.

The behavioral methods are as follows:

a. Inspiration implies moving students using intriguing facts, curiosity, fun ideas, and odd examples.
b. Engagement implies the accommodation students with different backgrounds by adapting lesson plans.
c. Entertainment implies injecting joyful activities in lessons.
d. Bonding implies establishing a close and empathetic relationship between students and their teacher.

We distributed a survey to students that consisted of three questions scored on a Likert Scale of 1 (strongly disagree) to 5 (strongly agree):

a. Post-class, I retain a clear grasp of the mathematical concepts and recall my teacher's instructions.
b. The mathematics lessons in my class are engaging, and my teacher motivates me to deepen my knowledge of mathematics.
c. I am in favor of future mathematics courses being conducted in a comparable manner.

E. Preprocessing

The original dataset contains 72 teacher surveys. However, the amount of student responses was inconsistent between teachers to a large degree. For example, some teachers received many responses while other teachers received very few. As such, certain teacher's associated student surveys were ignored to mitigate the effects of low sample size.

Leaving these data samples as is would severely bias the results as certain teachers who received few student responses could potentially enjoy a very high score which may not be representative of their entire class. We will explain how we resolved this issue shortly.

In this research, student scores were pre-processed by first calculating the average of the three student survey question scores. These averaged scores were then further averaged across each teacher's class to create a single value, known as the Student Evaluation Index (SEI). As a result, the decision tree algorithm's output class is now represented by this unified SEI value. This dual averaging approach simplifies the decision tree's task and can help eliminate outliers caused by student bias or other noise. However, it also reduces the dataset's granularity, potentially omitting important information that could impact accuracy.

For teachers whose student scores were dropped, a proxy for SEI is introduced which is the average score of the teacher's own technique and behavior scores. While this allows us to retain all the teachers which increase the number of data samples for the decision tree models, it can potentially generate incorrect predictions and feature importance rankings. Further discussion about the effects can be found in the next section.

A duplicate dataset is generated where the SEI is grouped into 0.25 intervals to facilitate the use of the classification decision tree. For instance, an SEI of 4.3 is rounded to 4.25, the nearest 0.25 interval. The classification decision tree will then use 4.25 as the output label for this data point. For the regression decision tree, the SEI remains unchanged.

Due to the scoring ranges of all input features being on a Likert scale of 1 to 5, there is no need to normalize the data. However, copies of the data for both classification and regression datasets are made that is standardized, removing the mean and scaling with the unit variance for use with PCA. Furthermore, datasets which contain 95% of the variance, as determined by using PCA, are generated. After the teacher and student survey preprocessing, this resulted in datasets containing 72 teacher feature inputs and 72 SEI outputs. Further discussion on the results is next.

III. Results and Discussion

Python3 and Scikit-Learn were utilized to implement the PCA and decision tree algorithms, while Pydotplus and Matplotlib were employed to visualize the resulting tree.

A. Classification Tree

To find the best classification tree, the dataset is split into 70% and 80% training data leaving 30%, and 20% of the data for testing, respectively. Two additional datasets split with these ratios are generated, one normally scaled, and the other contains features that PCA determines contains >95% of the variance. GridSearchCV was employed to identify the optimal hyperparameters for each classification tree based on various data splits. The GridSearchCV function utilized 5-fold cross-validation. We observed that the best performer tree was the one trained on the 80-20 PCA data, with a minimum of 1 sample per leaf and a maximum tree depth of 4. Impurity and information gain were measured using the Gini index metric. The classifier generated a feature importance ranking, which could be directly extracted. Results are detailed in section III.

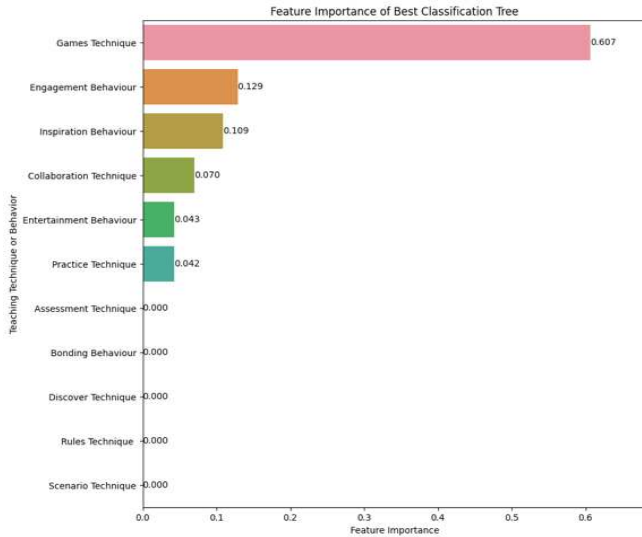| | Classification | | Regression | |
	Accuracy		MSE & R2	
Scoring				
Dataset Split	70-30	80-20	70-30	80-20
Original	41%	27%	0.244	0.392
			0.271	-0.098
Scaled	23%	20%	0.256	0.443
			0.238	-0.238
PCA	41%	67%	0.114	0.192
			0.661	0.464

Fig. 2 Feature importance ranking of the best performing classification tree, PCA 80-20. Note, after PCA, 5 additional features were deemed to not affect decision tree prediction ability
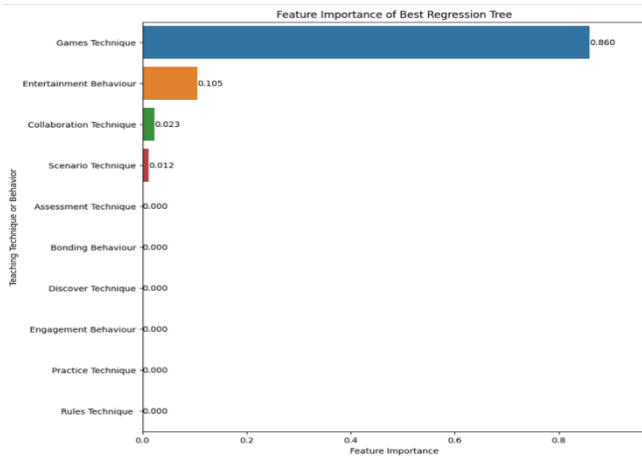

Fig. 3 Feature importance ranking of the best performing regression tree, PCA 70-30. Note, after PCA, 6 further features are not used at all in the tree and the decision largely rests with the Gamification technique

B. Regression Tree

Once again, to find the best regression tree, the dataset is split into 70% and 80% training data leaving 30%, and 20% of the data for testing. Two additional datasets are created from these splits, one normally scaled, and the other contains features that PCA determines contains >95% of the variance. Using 5-fold cross-validation, GridSearchCV identified the optimal regression tree, which utilized a 70-30 PCA split. This configuration resulted in a maximum tree depth of 4 and

a minimum of 3 samples per leaf. The Mean Squared Error (MSE) was used as the information gain metric for the regression tree, while the R-Squared (R2) score measured its performance. The classifier also generated a feature importance ranking, which could be directly extracted. Results are shown in the next section.

C. Classification Tree's Performance

The best performing classification tree trained on the 80-20 PCA dataset had an accuracy of 67%. PCA determined that the following features contain low variance: Visualization and Language. The feature rankings derived from the tree's structure yielded the following features as the most influential on SEI: Gamification, Engagement, Inspiration, Collaboration and Entertainment. Results can be found in Table I and feature importance can be found in Fig. 2.

D. Regression Tree's Performance

The best performing regression tree trained on 70-30 PCA data yielded an R2 score of 0.661 and a MSE of 0.114. Consequently, on average, the regression prediction was 0.338 points away from the "correct" Student Evaluation Index score. PCA determined that the following features did not contribute much information: Visualization, Language, and Inspiration. The top 4 most important features are: Gamification, Entertainment, Collaboration, and Case Studies. Further techniques and behaviors were deemed to add no extra information to the tree. Results can be found in Table I and feature importance can be found in Fig. 3.

IV. CONCLUSION

The accuracy of the classification tree when it came to predicting the SEI was only 67%. The subpar precision may be due to the coarse categorization of SEI, as detailed in Section III. Grouping the SEI into quarter-point increments could result in adjacent predictions being classified as incorrect, even though they are nearly correct. Qualitatively, this can still be considered an exact prediction. Quantitatively, this is incorrect and as such, accuracy suffers. As such, there likely exists a subset of predictions that can be considered qualitatively accurate despite being quantitatively inaccurate.

Furthermore, the augmentation of student response data for teachers with extremely low response rates can affect the results as well. Due to the nature of the augmentation, averaging the teaching technique and behavior scores for that teacher, teachers with self-reported high scores will automatically have a high SEI. This may not be true of real-life student responses as they do not necessarily vary linearly with teacher scores.

The regression tree's accuracy performance was also mediocre with an R2 score of 0.661 and an MSE of 0.114. As mentioned before, the classification tree aggressively bins SEI scores into groups as opposed to the regression tree. However, this does not necessarily increase the accuracy of the regression tree as the predicted SEI values can still lie beyond what is considered quantitatively accurate.
The MSE of 0.114 points translates into an average SEI error of 0.338 points for any given prediction or about 1.5 intervals. Once again, this can be considered decently accurate qualitatively in the context of teacher evaluation.

It should be noted that the R2 score in the context of decision trees is not an absolute measure of accuracy. However, an R2 of 0.661 indicates that the regression tree can model this dataset decently.

A. Ranking Features based on Importance

Ranking features based on their importance gave intriguing insights (see Fig. 2 and Table II). Despite the dataset undergoing PCA, the best classification tree only used 6 of the remaining 11 features. Additionally, *Language*, using the student's native language, can be logically deduced to be one of the important factors in student satisfaction and understanding. However, the PCA deems it to have little effect on SEI. The classification tree ranks Gamification as the most important feature by a large margin. Using games to teach greatly influences the SEI with a score of 0.607.

TABLE II
FEATURE IMPORTANCE SCORES OF THE BEST DATASETS AND DATASET SPLITS FOR EACH TREE

	Classification		Regression	
Split Ratio	70-30	80-20	70-30	80-20
Best Dataset	PCA	PCA	PCA	PCA
Techniques				
Visualization	0.12	0.042	0.0	0.0
Practice	0.0	0.0	0.0	0.0
Math Rules	0.0	0.0	0.0	0.0
Gamification	0.588	0.607	0.86	1.0
Collaboration	0.314	0.07	0.023	1.0
Discovery	0.0	0.0	0.0	0.0
Case Studies	0.0	0.0	0.0	0.0
Assessment	0.0	0.0	0.012	0.0
Language	0.0	0.0	0.0	0.0
Behaviors				
Inspiration	0.095	0.109	0.0	0.0
Engagement	0.098	0.129	0.0	0.0
Entertainment	0.0	0.043	0.105	0.0
Bonding	0.0	0.0	0.0	0.0

This is slightly more than quadruple the importance of the next feature, Engagement, at 0.129. Of less significant importance are Inspiration, Collaboration, Entertainment, and Practice. By extracting the ranking of features derived from the best classification tree, we may conclude that that teachers can achieve the quickest improvement in SEI by using the games to teach and actively engaging with students. However, due to the mediocre accuracy of the classification tree, the suggested priorities for areas of performance could be suboptimal, out of order or of incorrect magnitude.

The best regression tree's feature importance rankings result in interesting conclusions as well and can be seen in Fig. 3 and Table II. Due to using the PCA dataset the importance of each feature is more concentrated like the classification tree. Despite removing 3 features with PCA, the best regression tree only deems 4 of the remaining 10 features to contain any useful information.

The regression tree's feature importance ranking is heavily skewed, like the classification tree, towards Gamification at a score of 0.86. This means that 86% of the information gained in the regression tree can be attributed to the Gamification feature. This is more than 8 times more important than the next feature.

This strongly suggests that educators may need to focus on educational games to improve SEI. Remember, SEI is not only student performance, but student satisfaction as well. As such, Games of all types are inherently a form of fun or entertainment. It intuitively follows that they are popular with almost everyone, no matter the age.

In summary, the research conducted suggests that educators can improve their usage of games in their teaching as well as improving their behavioral techniques to improve the performance and satisfaction of their students the quickest.

B. Limitations

One potential limitation with respect to predicted SEI is the two rounds of averaging during the data preprocessing. This causes a lot of potentially useful trends in data to be simplified or hidden. Further research with not augmented student data could potentially yield different results as well. Another limitation is the relatively small number of survey participants. The performance of both trees could be more satisfactory if given more teacher-student samples to draw from. The survey format was intended for quick completion to avoid loss of interest or diligence by the participants. Specifically for educators, certain teaching or behavioral methods could potentially be misunderstood as something else due to the short explanations.

Finally, the choice of decision tree addresses both research questions put forth in this paper. However, small changes in training data can radically affect tree structure which informs feature importance ranking as can be seen with the use of PCA. Additionally, decision trees are suboptimal for regressing continuous values and are prone to overfitting which can affect the generalizability of SEI prediction.

C. Summary of Findings

The research conducted in this paper demonstrates the ability to predict teacher performance based on their utilization of teaching techniques and different behavioral skills. Additionally, it explores and demonstrates the possibility of discovering which techniques and behaviors can be prioritized for better teacher training.

This second aim is crucial for many school systems, especially in Dubai, to better prepare students for the future. It is imperative that modern classrooms should take advantage of modern technology to optimize the learning experience for both teachers and students alike. Through the optimization of teacher performance and the enhancement of their experience, students can anticipate an increase in their own satisfaction as well as improved uptake and retention of knowledge.

D. Future Work

Some potential avenues for future investigation are different decision tree variations such as ensemble learning with Random Forests or gradient boosting. This will retain the inherent feature importance ranking functionality while potentially improving on SEI prediction.

Moreover, the influence of geographical, demographic, and cultural elements may extend to both the forecasting of Student Evaluation of Instruction (SEI) and the identification of the most effective pedagogical strategies and actions. Conducting a comparable investigation in a different nation might confirm universally accepted educational standards or

reveal tactics and practices that are significant within a specific regional context.

Finally, the current feature importance implementation returns a ranking for the entire population of teachers. Potential future work can include personalized feature importance ranking which would exclude features that the teacher already uses to a high degree.

REFERENCES

[1] "The world needs almost 69 million new teachers to reach the 2030 education goals," UNESCO Institute for Statistics, Oct. 2016, [Online]. Available: https://uis.unesco.org/en/files/fs39-world-needs-almost-69-million-new-teachers-reach-2030-education-goals-2016-en-pdf

[2] A. M. Jarrah, "The Challenges Faced by Pre-Service Mathematics Teachers during their Teaching Practice in the UAE: Implications for Teacher Education Programs," International Journal of Learning, Teaching and Educational Research, vol. 19, no. 7, pp. 23–34, Jul. 2020, doi: 10.26803/ijlter.19.7.2.

[3] F. Kayan Fadlelmula, "Enablers and obstacles in teaching and learning of mathematics," LUMAT: International Journal on Math, Science and Technology Education, vol. 10, no. 2, Jun. 2022, doi:10.31129/lumat.10.2.1627.

[4] L. Darling-Hammond, M. Hyler, and M. Gardner, "Effective Teacher Professional Development," Learning Policy Institute, Jun. 2017. doi:10.54300/122.311.

[5] E. Alpaydin, Introduction to Machine Learning, third edition, 3rd ed. MIT Press, 2014.

[6] L. Chen, P. Chen, and Z. Lin, "Artificial Intelligence in Education: A Review," IEEE Access, vol. 8, pp. 75264–75278, 2020, doi:10.1109/access.2020.2988510.

[7] M. Chassignol, A. Khoroshavin, A. Klimova, and A. Bilyatdinova, "Artificial Intelligence trends in education: a narrative overview," Procedia Computer Science, vol. 136, pp. 16–24, 2018, doi:10.1016/j.procs.2018.08.233.

[8] M. Wollowski et al., "A Survey of Current Practice and Teaching of AI," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 30, no. 1, Mar. 2016, doi: 10.1609/aaai.v30i1.9857.

[9] E. M. Ganyaupfu, "Teaching methods and students' academic performance," International Journal of Humanities and Social Science Invention, vol. 2, no. 9, Sep. 2013.

[10] R. Tripathi, "Importance and Improvements in Teaching-Learning process through Effective Evaluation Methodologies," ESSENCE International Journal for Environmental Rehabilitation and Conservation, vol. 9, no. 2, pp. 7–16, Dec. 2018, doi:10.31786/09756272.18.9.2.202.

[11] E. Akdemir and C. Özçelik, "The Investigation of the Attitudes of Teachers towards Using Student Centered Teaching Methods and Techniques," Universal Journal of Educational Research, vol. 7, no. 4, pp. 1147–1153, Apr. 2019, doi: 10.13189/ujer.2019.070427.

[12] S. Petrina, "Advanced Teaching Methods for the Technology Classroom," 2007, doi: 10.4018/978-1-59904-337-1.

[13] P. Sokkhey and T. Okazaki, "Comparative Study of Prediction Models for High School Student Performance in Mathematics," IEIE Transactions on Smart Processing & Computing, vol. 8, no. 5, pp. 394–404, Oct. 2019, doi: 10.5573/ieiespc.2019.8.5.394.

[14] L. Dinh, J. Sohl-Dickstein, and S. Bengio, "Density estimation using Real NVP," arXiv (Cornell University), Jan. 2016, https://arxiv.org/abs/1605.08803

[15] R. Russell, Machine learning step-by-step guide to implement machine learning algorithms with Python. 2010.

[16] A. M.O., O. A.O., and W. W.F., "Teachers' Performance Evaluation in Higher Educational Institution using Data Mining Technique,"

[17] M. Ünal, "Preferences of Teaching Methods and Techniques in Mathematics with Reasons," Universal Journal of Educational Research, vol. 5, no. 2, pp. 194–202, Dec. 2017, doi:10.13189/ujer.2017.050204.

[18] M. Joubert, "Using digital technologies in mathematics teaching: developing an understanding of the landscape using three 'grand challenge' themes," Educational Studies in Mathematics, vol. 82, no. 3, pp. 341–359, Sep. 2012, doi: 10.1007/s10649-012-9430-x.

[19] Y. Wardat, M. Tashtoush, R. Alali, and S. Saleh, "Artificial intelligence in education: mathematics teachers' perspectives, practices and challenges," Iraqi Journal for Computer Science and Mathematics, vol. 5, no. 1, pp. 60–77, Jan. 2024, doi:10.52866/ijcsm.2024.05.01.004.

[20] J. Aliyu, S. Osman, M. F. Daud, and J. A. Kumar, "Mathematics teachers' pedagogy through technology: a systematic literature review," International Journal of Learning Teaching and Educational Research, vol. 20, no. 1, pp. 323–341, Jan. 2021, doi:10.26803/ijlter.20.1.18.

[21] S. A. Pise, "Transforming mathematics education with technology," An International Peer Reviewed Scholarly Research Journal for Interdisciplinary Studies, p. 17, 2013.

[22] L. P. Alonso, "Learning Process Analysis using Machine Learning Techniques," Mar. 2019, doi: 10.31219/osf.io/srhz9.

[23] Z. Yujiao, L. W. Ang, S. Shaomin, and S. Palaniappan, "Dropout Prediction Model for College Students in MOOCs Based on Weighted Multi-feature and SVM," Journal of Informatics and Web Engineering, vol. 2, no. 2, pp. 29–42, Sep. 2023, doi: 10.33093/jiwe.2023.2.2.3.

[24] T. Ahmed Khan, R. Sadiq, Z. Shahid, M. M. Alam, and M. Mohd Su'ud, "Sentiment Analysis using Support Vector Machine and Random Forest," Journal of Informatics and Web Engineering, vol. 3, no. 1, pp. 67–75, Feb. 2024, doi: 10.33093/jiwe.2024.3.1.5.

[25] S. M. Wajid, T. Javed, and M. Mohd Su'ud, "Ensemble Learning-Powered URL Phishing Detection: A Performance Driven Approach," Journal of Informatics and Web Engineering, vol. 3, no. 2, pp. 134–145, Jun. 2024, doi: 10.33093/jiwe.2024.3.2.10.

[26] T. S. Tajamul and K. Aman, "Unveiling the Efficacy of AI-based Algorithms in Phishing Attack Detection," Journal of Informatics and Web Engineering, vol. 3, no. 2, pp. 116–133, Jun. 2024, doi:10.33093/jiwe.2024.3.2.9.

[27] H. Ng, M. S. Jalani, T. T. V. Yap, and V. T. Goh, "Performance of Sentiment Classification on Tweets of Clothing Brands," Journal of Informatics and Web Engineering, vol. 1, no. 1, pp. 16–22, Mar. 2022, doi: 10.33093/jiwe.2022.1.1.2.

[28] J. S. A. Tan, Z. Che Embi, and N. Hashim, "Comparison of Machine Learning Methods for Calories Burn Prediction," Journal of Informatics and Web Engineering, vol. 3, no. 1, pp. 182–191, Feb. 2024, doi:10.33093/jiwe.2024.3.1.12.

[29] S. T. Lim, J. Y. Yuan, K. W. Khaw, and X. Chew, "Predicting Travel Insurance Purchases in an Insurance Firm through Machine Learning Methods after COVID-19," Journal of Informatics and Web Engineering, vol. 2, no. 2, pp. 43–58, Sep. 2023, doi:10.33093/jiwe.2023.2.2.4.

[30] J. O. Victor, X. Chew, K. W. Khaw, and M. H. Lee, "A Cost-Based Dual ConvNet-Attention Transfer Learning Model for ECG Heartbeat Classification," Journal of Informatics and Web Engineering, vol. 2, no. 2, pp. 90–110, Sep. 2023, doi: 10.33093/jiwe.2023.2.2.7.

[31] T.-J. Law, C.-Y. Ting, H. Ng, H.-N. Goh, and A. Quek, "Ensemble-SMOTE: Mitigating Class Imbalance in Graduate on Time Detection," Journal of Informatics and Web Engineering, vol. 3, no. 2, pp. 229–250, Jun. 2024, doi: 10.33093/jiwe.2024.3.2.17.

[32] C. Susaie, C.-K. Tan, and P.-Y. Goh, "Learning Experience with LearnwithEmma," Journal of Informatics and Web Engineering, vol. 1, no. 2, pp. 30–44, Sep. 2022, doi: 10.33093/jiwe.2022.1.2.3.

International Journal of Applied Information Systems, vol. 10, no. 7, pp. 10–15, Mar. 2016, doi: 10.5120/ijais2016451524.