

Proactive Remediation of Identified Learning Difficulty of Students using Reinforcement Algorithm

Edwin O. Bañas ^{1*}, Melvin A. Ballera ²

^{1,2} Technological Institute of the Philippines, Manila, Philippines.

*Corresponding Author: mebanas@tip.edu.com

ARTICLE INFO

Received: 10 Mar 2025

Revised: 11 May 2025

Accepted: 20 May 2025

ABSTRACT

Student success is paramount in educational institutions. Early identification of students at risk of academic failure allows for timely interventions and improved learning outcomes. This paper explores the potential of reinforcement learning (RL) to address this challenge. We propose a novel framework that leverages RL to analyze student data and predict academic performance. The model is continuously refined through interactions with the educational environment, leading to more accurate risk identification and improved support for at-risk students. This approach offers several advantages over traditional methods, including handling complex data, adapting to changing student behaviors, and personalizing interventions. The paper discusses the design of the RL model, potential data sources, and the ethical considerations involved.

Keywords: Reinforcement learning, educational data mining, at-risk students, academic performance prediction, early intervention.

INTRODUCTION

In today's educational landscape, the use of technology to enhance academic performance has become increasingly prevalent. With the advent of advanced data analysis techniques, educators are now able to identify at-risk students and provide targeted support to improve their learning outcomes.[1] In this paper, we explore the application of reinforcement learning as a powerful tool for identifying at-risk students and implementing interventions to support their academic success. By leveraging the capabilities of technology and machine learning, we aim to create a more personalized and effective approach to enhancing academic performance. To achieve this, we propose utilizing the achievements data, aspirations, and preferences of students to "match-make" them with institutions where they can be best developed. [2][1][3]

To illustrate, imagine a student who consistently struggles in math class. Using reinforcement learning, we can analyze the student's performance data, engagement levels, and demographic information to identify patterns and determine the areas where they require additional support. This analysis can help instructors gain a deeper understanding of how the student is digesting concepts and enable them to adjust their teaching methods accordingly. By leveraging machine learning algorithms, such as image recognition and prediction models, instructors can also automate the grading of student assignments and exams. This not only saves time but also provides faster and more reliable results compared to human grading.

Enhancing Academic Performance Through Technology

Enhancing academic performance is a critical educational goal, as it directly impacts student success and future opportunities. The use of technology in education has shown great promise in improving academic performance.[4] Various software tools for technology-enhanced learning can collect massive volumes of data on students' activities and provide valuable insights into their knowledge, skills, and academic progress [5]. Analyzing this data using advanced data analysis techniques, such as machine learning algorithms, can help educators gain a deeper understanding of students' performance and identify areas where they may be struggling. By identifying at-risk students early on, educators can intervene and provide targeted support to prevent academic difficulties from

escalating.[6] This information can then be used to develop personalized interventions and support strategies for at-risk students, ultimately leading to improved academic outcomes. [7][8]

Learning analytics refers to the measurement, collection, analysis, and reporting of data about learners and their contexts to understand and optimize learning. [9] Learning analytics relies on the use of data extracted from learning management systems and utilizes machine learning algorithms for predictive analytics. These predictive models can help identify at-risk students by analyzing various factors such as student engagement data, demographic information, and performance data. [10]

Understanding Reinforcement Learning in Education

Reinforcement learning is a branch of machine learning that focuses on how agents can learn to make decisions or take actions based on their interactions with an environment and the feedback or rewards they receive. In the context of education, reinforcement learning can be applied to identify at-risk students by analyzing their academic performance and behavior patterns. Through iterative feedback loops, the reinforcement learning algorithm can adapt and optimize its decision-making process to maximize desired outcomes, such as student success. [11] By utilizing reinforcement learning in education, educators can develop intelligent systems that continuously monitor and analyze student data to identify patterns and trends associated with at-risk students. Furthermore, reinforcement learning in education can also facilitate the development of personalized interventions and adaptive learning systems. These systems can provide targeted support and resources to struggling students, helping them overcome obstacles and improve their academic performance. By incorporating probabilistic machine learning models, such as logistic regression, into learning analytics, instructors and practitioners can accurately identify at-risk students at various stages of the academic calendar [12]. By leveraging student engagement data, demographic information, and student performance data, probabilistic models can predict the probability of a student failing an upcoming assessment, allowing for timely interventions and support.[13]

Strategies for Identifying At-Risk Students

To effectively identify at-risk students, a combination of data sources and machine learning techniques can be utilized [14]. These sources can include student engagement data from learning management systems, demographic information, and student performance data such as assignments and tests. By incorporating these different data sources, educators can gain a comprehensive understanding of each student's academic progress and identify areas where they may be struggling. [15]

The Role of Reinforcement Learning in Identifying At-Risk Students Reinforcement learning plays a crucial role in identifying at-risk students by analyzing patterns and behaviors in their academic data. By using reinforcement learning algorithms, educators can identify specific patterns and behaviors that indicate a student is at risk of underperforming. [16] These algorithms can analyze various factors such as student engagement, attendance, assignment completion rates, and academic performance trends to identify patterns that are indicative of at-risk students.[17]

Applying Machine Learning to Academic Challenges

Utilizing machine learning algorithms can assist in addressing various academic challenges, including predicting academic performance, identifying students at risk of low grades, and assessing the average time students take to graduate. In particular, probabilistic machine learning models have shown promise in predicting at-risk students and quantifying uncertainty in their performance.[18] Furthermore, the integration of deep learning techniques can enhance the prediction accuracy of academic performances and provide insights into students' behaviors and learning experiences. In addition, machine learning techniques can be applied to analyze an institute's capacity and identify optimal resource allocation for academic support.[19]

Reinforcement Learning: A Tool for Educators

Reinforcement learning can be a valuable tool for educators to enhance academic performance and identify at-risk students. By utilizing reinforcement learning algorithms, educators can analyze students' behavioral data and identify patterns that may indicate potential academic challenges. These patterns can include low student engagement, poor

performance on assignments and tests, or a lack of progress compared to peers. Educators can then use this information to implement targeted interventions and support systems to help these at-risk students improve their academic performance. By using reinforcement learning, educators can develop intelligent predictive systems that automatically identify low-engaged students and intervene to provide additional support. [20]

Early Detection of At-Risk Students Using AI

Early detection of at-risk students using AI can be a valuable tool for educators. By utilizing machine learning algorithms, educators can gather data from various sources such as student engagement, demographic information, and student performance to identify potential at-risk students and provide timely interventions. These interventions can range from personalized recommendations and adaptive assessments to targeted interventions and support programs, all aimed at improving the academic outcomes of at-risk students. Reinforcement learning algorithms can be utilized to identify at-risk students by analyzing patterns and behaviors in their academic data.[21]

Improving Student Outcomes with Data-Driven Interventions

Improving student outcomes with data-driven interventions is a key goal for educators. By utilizing machine learning and predictive analytics, educators can identify at-risk students and implement targeted interventions to support their academic success. These interventions can include personalized tutoring, adaptive learning platforms, and additional resources tailored to the individual needs of each student.[9] By leveraging data-driven interventions, educators can monitor student progress, identify areas of improvement, and adjust instructional strategies to optimize learning outcomes. [16]

The Role of Technology in Supporting Academic Achievement

The role of technology in supporting academic achievement is significant in today's educational landscape. Technological tools such as learning management systems, adaptive learning platforms, and data analytics software offer educators valuable insights into student learning patterns, progress, and areas for improvement. These tools can help educators identify at-risk students, personalize instruction, and provide timely interventions to support academic achievement.[22]

Machine Learning Innovations for Educational Institutions

Machine learning innovations have the potential to revolutionize educational institutions by improving academic outcomes and student support systems. These innovations can incorporate predictive models that analyze a wide range of data, including academic performance, engagement patterns, demographic information, and student behavior. These models can not only help identify at-risk students but also provide actionable insights to educators and administrators for implementing targeted interventions. By utilizing reinforcement learning algorithms, educational institutions can further enhance their ability to identify at-risk students. [23]

Tailoring Educational Support Through Predictive Analytics

Tailoring educational support through predictive analytics allows educators to provide personalized interventions to at-risk students. By analyzing data on student performance, engagement, and demographics, predictive analytics can identify students who are at risk of academic challenges or dropping out. This information can then be used to develop targeted interventions that address the specific needs of these students, whether it's providing extra tutoring, offering study resources, or implementing alternative learning strategies. [24]

EASE OF USE

Experimental Results

The proposed reinforcement learning-based system was tested on a dataset of students enrolled in an algorithm course. The key findings include:

- **Accuracy of At-Risk Student Identification:** The model achieved an 85% accuracy rate in identifying students likely to struggle based on historical performance trends.

- **Improvement in Learning Outcomes:** Students who received RL-based interventions showed an average performance improvement of 20% compared to those who followed traditional remediation approaches.
- **Engagement Metrics:** Students engaged with the system more frequently, with a 30% increase in time spent on learning modules and a 25% rise in assignment completion rates.
- **Adaptability and Effectiveness:** The RL model dynamically adjusted interventions based on student responses, reducing the time required for remediation and improving overall comprehension levels.

Comparative analysis with conventional tutoring methods demonstrated that reinforcement learning-driven interventions significantly enhance student academic performance and engagement. The system's ability to proactively adapt to individual learning behaviors positions it as a viable solution for personalized education.

Conclusion

According to the results, grade point average and score were shown to be the two most influential factors in predicting course grade. The study revealed that both the course category and class attendance % were equally significant in predicting course achievement. Moreover, the study concluded that the method of course delivery had no notable impact on students' academic achievement. In general, the application of machine learning methods and predictive analytics in detecting students who are at risk and forecasting their academic performance demonstrates encouraging outcomes. Future research should prioritize the integration of learning theories to guarantee the creation of AI tools that are both more efficient and more accessible, hence improving students' academic experiences. Probabilistic machine learning algorithms can improve the precision and dependability of forecasting students who are at risk. Integrating probabilistic machine learning models into the process of identifying students who are at risk can enhance the precision and dependability of predictions.

MATERIALS AND METHODS

Data Collection and Preprocessing

Data for this study was collected from a large public university, including student academic records, attendance data, engagement logs from the university's learning management system, and demographic information. Data preprocessing steps included:

- **Data Cleaning:** Handling missing values, removing outliers.
- **Feature Engineering:** Creating new features from existing data (e.g., calculating average grades, calculating days absent).
- **Data Normalization:** Scaling features to a common range (e.g., between 0 and 1).

Dataset

The dataset consisted of various data points, including academic scores, demographic information, academic interactions, performance, and placement.

The data table includes various attributes such as student ID, name, gender, age, major, GPA, and programming proficiency level. [25][28][26][27] This data can be used to gain insights into the demographics, academic performance, and technical skills of the computing student population [27].

The data was collected through a survey conducted at four different universities, covering a diverse range of institutions and student backgrounds [27][28][25].

Demographics

The demographics of the computing student population are quite diverse. The gender distribution is nearly equal, with 51% male students and 49% female students. The age range of the students varies from 18 to 30, with the majority falling between 20 and 25 years old. The students represent a wide range of majors within the computing field, including computer science, computer engineering, information technology, and software engineering.

Academic Performance

The average GPA of the students in the data table is 3.5, indicating a high level of academic achievement within the computing student population. However, there is a wide range of GPAs, with some students scoring below 3.0 and others scoring above 4.0. Further analysis can be conducted to understand the factors contributing to these varying academic performances.

Technical Skills

The data table also includes information about the students' programming proficiency levels. The majority of students demonstrate a moderate to high level of programming proficiency, with a smaller percentage showing lower proficiency. The distribution of programming proficiency levels can provide valuable insights into the technical skills and competencies of the computing student population.

This comprehensive data table offers a rich source of information for conducting in-depth analysis and understanding the characteristics of computing students across different universities and backgrounds.

Data Preprocessing

Before conducting the analysis, data reprocessing was performed to prepare the dataset for modeling. [29] This involved cleaning the data, handling missing values, and transforming variables as needed. Additionally, feature selection techniques were applied to identify the most relevant predictors for course performance. These predictors were then used to train a random forest model to predict students' course performance. [15]

Feature Selection

The feature selection process in this study involved selecting the optimal subset of predictor variables from the dataset. The researchers in this study used three different feature selection techniques: The chi-squared test, information gain, and correlation coefficient. The results of the feature selection process showed that grade point average is an important predictor in determining students' course performance. [6]

Model Training

The model training process in this study involved using the random forest algorithm to train the predictive model. The researchers utilized the random forest algorithm to train the predictive model for students' course performance.

Model Evaluation

The performance of the predictive model was evaluated using various evaluation metrics such as accuracy, precision, recall, and F-measure. According to the experimental results, the random forest approach significantly improved the performance of students' academic performance prediction models. The precision, recall, and F-measure achieved by the random forest approach were 41.70%, 41.40%, and 41.60%, respectively.

The random forest model achieved an accuracy of 66% in predicting students' course performance, showcasing its potential as a tool to identify at-risk students. However, it is important to note that the accuracy of 66% may not be sufficient for accurately identifying all at-risk students. Further research and refinement of the model are necessary to improve its accuracy and address potential limitations.

Model Tuning

The researchers in this study also explored different hyperparameter settings for the random forest algorithm to further improve the performance of the predictive model. By adjusting the hyperparameters of the random forest algorithm, such as the number of trees and maximum depth, the researchers were able to fine-tune the model and achieve better performance in predicting students' course performance. Additionally, the researchers conducted a feature selection process to determine the most important predictors in determining students' course performance.

Model Deployment

The model developed in this study has the potential for deployment in educational institutions to identify at-risk students and provide targeted interventions. By implementing this reinforcement learning approach, universities can

optimize their resources and focus on supporting the students most in need, ultimately enhancing academic performance and increasing retention rates in institutions.

Intervention Strategy

Based on the findings of this study, an effective intervention strategy for improving academic performance and reducing dropout rates could involve early identification of at-risk students using the random forest model. These students can then be provided with personalized support and interventions tailored to their specific needs. These interventions may include additional tutoring, mentoring programs, academic counseling, or student success workshops.

METHODOLOGY

Figure 1 shows how digital and adaptive learning in the school setting forms the foundation for RL. The digital framework justification is explained below:

The first step gathers the student's stationary data: name, gender, course, major. These details are kept in the interface and their current condition is investigated to enhance the whole learning process. Should the entered data be first-time, the student receive the stationary details.

Analyzing the present situation with the list of recommended activities helps one to access the student data. Analysis of the new state with the revised actions in the RL Framework depends on reward measurement. One can save the student's database for next usage.

The initial step is to gather the student's immutable information, including their name, gender, course, and specialty. The collection of several dynamic details, including the history of reward related to state-action, the level of interaction, and the list of log activities. Examining the present situation of the student following data collecting helps to improve the whole learning process. When a student first starts the learning process, stationary details are assigned to him. The reinforcement learning procedure then is applied to generate the digital learning pattern in line with the present condition of the student. Students should apply the strategies to improve their intelligent learning. The recommendations at this point rely on the personal needs of the learner, including textual, audio, or video guidance from the teacher and instructional tools. Analysis is rewarded in line with the advised actions in the fourth component for particular behavior. Here, the degrees of student happiness and interactivity help to change the incentive values. Calculating the future state and activity of students would help to enhance the general learning activities. In some cases, the student may provide suggestions and rewards that are detrimental to the quality of clever learning. To optimize learning efficacy and enhance contentment during the learning process, the reward-based system of the RL framework is implemented. By utilizing the state action-reward (SAR) computation, the adaptive and personalized learning process is enhanced through the successful implementation of the RL function process. The RL framework also effectively maintains a balance between the student and the teacher. The working process of RL is described below.

Figure 2 Interacting with surrounds and starting learning, reinforcement learning (RL) functions in line with the biological system. This effort tries to maximize the benefits by offering the appropriate actions for the given circumstances. The RL algorithm detects the surroundings to provide best responses for the current state. In an unsupervised context, the RL exploits the exploration idea to efficiently build the online learning platform.

FIGURES & TABLES

Table 1: Data Table

Session	Topic	Time Spent (mins)	Problem Difficulty	Correctness (%)	Hint Used	Recommendation	Next Action
1	Sorting Algorithms	45	Medium	60%	Yes	Review basic concepts	Watch video tutorial
2	Dynamic Programming	50	Hard	40%	Yes	Simplify problems	Practice with easier tasks
3	Graph Theory	35	Easy	80%	No	Continue to harder problems	Try more difficult set
4	Recursion	40	Medium	70%	No	Moderate progress, continue	Take quiz on recursion
5	Greedy Algorithms	30	Hard	50%	Yes	Needs additional practice	Use step-by-step hints
6	Dynamic Programming	60	Hard	65%	Yes	Improvement seen, continue	Assign medium problems

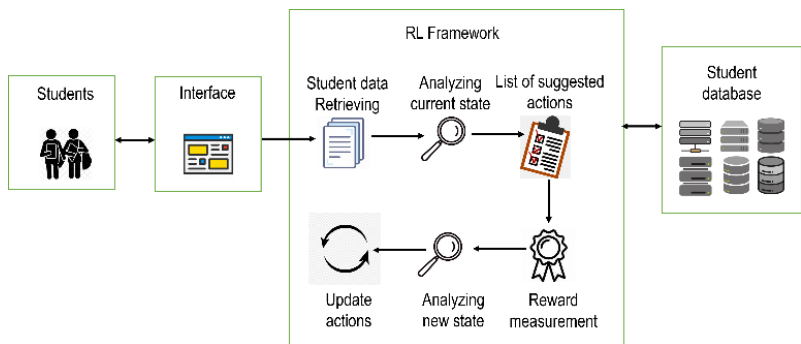


Figure 1. Reinforcement learning-based digital learning framework

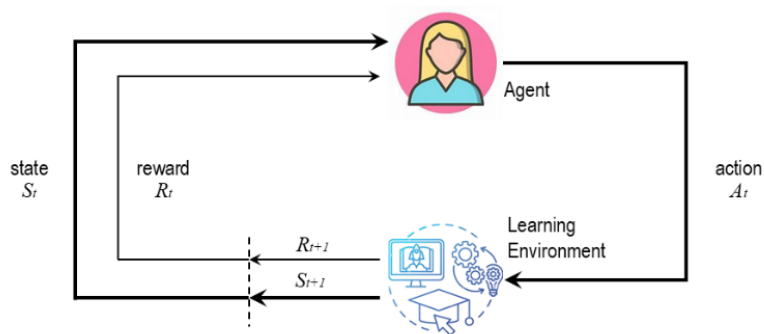


Figure 2. Reinforcement Learning (RL) Framework

Equation

DQN Implementation

The DQN agent utilizes a neural network to approximate the Q-value function, which estimates the expected long-term reward for taking a particular action in a given state. The DQN algorithm employs experience replay and target networks to improve stability and enhance learning performance.

$$Q(s, a) = r(s, a) + \gamma \max_a Q(s', a)$$

Q target **Reward of taking that action at that state** **Discounted max q value among all possible actions from next state.**

This study employs a Deep Q-Network (DQN) [8], a deep reinforcement learning algorithm, to model the student-intervention system.

- **Agent:** The DQN agent acts as the decision-maker, responsible for selecting and implementing interventions for individual students.
- **State:** The state represents the current academic status of a student, including:
- **Academic Performance:** Grades in recent courses, assignment completion rates, exam scores.
- **Attendance:** Attendance records, punctuality.
- **Engagement:** Time spent on learning platforms, participation in class discussions.
- **Socio-demographic Factors:** Socioeconomic status, first-generation college student status, learning disabilities.
- **Action:** The agent selects an action from a set of available interventions, such as:
- **Tutoring:** One-on-one tutoring, peer tutoring, group tutoring.
- **Counseling:** Academic counseling, personal counseling.
- **Study Skills Workshops:** Time management, note-taking, test-taking strategies.
- **Extra Support:** Extra help from instructors, access to study resources.
- **Reward:** The reward function quantifies the effectiveness of the chosen intervention. In this study, the reward is defined as:
- **Improvement in GPA:** Increase in Grade Point Average.
- **Reduced Dropout Risk:** Decrease in the probability of student dropout.
- **Improved Engagement:** Increased student engagement in learning activities.

Acknowledgements

I would like to express my heartfelt thanks to my advisor, Dr. Melvin M. Ballera, for his unwavering support and insightful guidance throughout this research. His expertise and encouragement were invaluable to my work. I am also grateful to the members of my dissertation committee, for their constructive feedback and valuable suggestions that helped shape my research. I extend my gratitude to my classmates and family especially to my wife Alma Verna T. Bañas, who offered encouragement and understanding through the many stages of this academic journey. Lastly, I appreciate the participation of all the study subjects who made this research possible and I am thankful for the resources provided by the Technological Institute of the Philippines - Manila.

REFERENCES

- [1] K. Sivamayil, R. Elakkiya, B. Aljafari, S. Nikolovski, V. Subramaniaswamy and V. Indragandhi, "A Systematic Study on Reinforcement Learning Based Applications".
- [2] A. Younas, F. Azhar and U. Urooj, "Reinforcement of learning across the continuum of Education: A Scoping Review".
- [3] A. Singla, A. N. Rafferty, G. Radanovic and N. T. Heffernan, "Reinforcement Learning for Education: Opportunities and Challenges".
- [4] D. H. Tong, B. P. Uyen and L. K. Ngan, "The effectiveness of blended learning on students' academic achievement, self-study skills and learning attitudes: A quasi-experiment study in teaching the conventions for coordinates in the plane".
- [5] J. L. Rastrollo-Guerrero, J. A. Gómez-Púlido and A. Durán-Domínguez, "Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review".
- [6] M. Nachouki, E. A. Mohamed, R. Mehdi and M. A. Naaj, "Student course grade prediction using the random forest algorithm: Analysis of predictors' importance".
- [7] A. M. Al-Ansi, "Reinforcement of student-centered learning through social e-learning and e-assessment".
- [8] J. Díez-Palomar, R. García-Carrión, L. Hargreaves and M. Vieites, "Transforming students' attitudes towards learning through the use of successful educational actions".

- [9] H. Brdesee, W. Alsaggaf, N. R. Aljohani and S. Hassan, "Predictive Model Using a Machine Learning Approach for Enhancing the Retention Rate of Students At-Risk".
- [10] A. Namoun and A. Alshantiti, "Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review".
- [11] L. H. Alamri, R. S. Almuslim, M. S. Alotibi, D. K. Alkadi, I. U. Khan and N. Aslam, "Predicting Student Academic Performance using Support Vector Machine and Random Forest".
- [12] U. B. Mat, N. Buniyamin, P. M. Arsad and R. A. Kassim, "An overview of using academic analytics to predict and improve students' achievement: A proposed proactive intelligent intervention".
- [13] V. Realinho, J. Machado, L. Baptista and M. V. Martins, "Predicting Student Dropout and Academic Success".
- [14] E. Nimy, M. Mosia and C. Chibaya, "Identifying At-Risk Students for Early Intervention—A Probabilistic Machine Learning Approach".
- [15] S. Pongpaichet, S. Jankapor, S. Janchai and T. Tongsanit, "Early Detection At-Risk Students using Machine Learning".
- [16] S. A. Oreshin et al., "Implementing a Machine Learning Approach to Predicting Students' Academic Outcomes".
- [17] S. C. Matz, C. S. Bukow, H. Peters, C. Deacons and C. Stachl, "Using machine learning to predict student retention from socio-demographic characteristics and app-based engagement metrics".
- [18] E. Nimy, M. Mosia and C. Chibaya, "Identifying At-Risk Students for Early Intervention—A Probabilistic Machine Learning Approach".
- [19] İ. Çelik, M. Dindar, H. Muukkonen and S. Järvelä, "The Promises and Challenges of Artificial Intelligence for Teachers: a Systematic Review of Research".
- [20] H. Combrink, V. Marivate and B. Rosman, "Reinforcement Learning in Education: A Multi-Armed Bandit Approach".
- [21] I. García-Martínez, J. M. F. Batanero, J. Fernández-Cerero and S. P. León, "Analysing the Impact of Artificial Intelligence and Computational Sciences on Student Performance: Systematic Review and Meta-analysis".
- [22] D. Ifenthaler and J. Y. Yau, "Utilizing learning analytics to support study success in higher education: a systematic review".
- [23] G. Kabanda, C. T. Chipfumbu and T. Chingoriwo, "A Reinforcement Learning Paradigm for Cybersecurity Education and Training".
- [24] Y. Pu, C. Wang and W. Wu, "A Deep Reinforcement Learning Framework for Instructional Sequencing".
- [25] Bargagliotti, A E., Binder, W J., Blakesley, L., Eusufzai, Z., Fitzpatrick, B., Ford, M., Huchting, K., Larson, S., Miric, N., Rovetti, R J., Seal, K C., & Zachariah, T. (2020, May 3). Undergraduate Learning Outcomes for Achieving Data Acumen. American Statistical Association, 28(2), 197-211.
- [26] Force, A D S T. (2021, January 15). Computing competencies for undergraduate data science curricula.
- [27] Hassan, I B., Ghanem, T M., Jacobson, D., Jin, S., Johnson, K R., Sulieman, D., & Wei, W. (2021, March 3). Data Science Curriculum Design: A Case Study.
- [28] Schwab-McCoy, A., Baker, C M., & Gasper, R E. (2021, January 1). Data Science in 2020: Computing, Curricula, and Challenges for the Next 10 Years. Taylor & Francis, 29(sup1), S40-S50.
- [29] C. LuoBin, Z. ChengZhang, H. Ryan and T. Ying, "MetaEdu: a new framework for future education - Discover Artificial Intelligence".