An Integrated Framework for an Educational Early Warning System with Mentor Matching

D. V. D. S. Abeysinghe
University of Moratuwa, Sri Lanka

M. S. D. Fernando
University of Moratuwa, Sri Lanka

Abstract

"Education is the key to success," one of the most heard motivational statements by all of us. People engage in education at different phases of our lives in various forms. Among them, university education plays a vital role in our academic and professional lives. During university education many undergraduates will face several challenges demanding from educational matters to socio-economic problems. In such situations, many undergraduates tend to abandon the degree programs halfway leaving them incomplete. Hence creating an Educational Early Warning Systems (EEWS) to predict and identify at-risk students in the early stages of the degree programs will improve the graduating ratio against the dropouts. Further, mentoring is another aspect in education where it can be used in undergraduate studies to address students individually. There exist many separate frameworks for EEWS and mentoring, but there exists a lacuna for an integrated framework for the two aspects. Having an integrated framework to identify at-risk undergraduates and matching the best matched mentor would be more impactful and effective for the universities to control dropouts. This study has proposed an integrated framework namely as "GRADGROOM" as a solution to the identified lacuna by extending EEWS framework with mentor matching which performs at-risk undergraduate prediction and mentor-mentee matching for them. Through two case studies at a local university, the study has concluded that a proper mentoring process conducted immediately after being identified as at-risk students will be highly beneficial to reshape their study patterns to align with the correct route of studying.

Keywords: at-risk student prediction, Educational Early Warning Systems (EEWS), mentor-mentee matching, virtual mentoring

Education will be the key for us to unlock the world. People will be engaged in education throughout their lives in divergent phases in numerous ways. In education, to bring it closer to the students, mentoring will play a significant role. The process of mentoring can be defined as the building of a relationship between a mentor and a mentee to boost the confidence and skills of a mentee. Mentoring has become an important part of education, which will enhance the reflective practice of education and improve the professional development of students (Sundli, 2007). With mentoring, rather than developing specific academic abilities, it will also focus on building a resilience character with confidence and good relationship handling (Martha, 2022).

Mentoring will have a lot of impactful benefits to both mentee and mentor (O'Connell, 2024). Through a proper mentoring program, a mentee will be able to identify the learning and skill areas to be improved through the constructive feedback they receive from mentors. Moreover, mentoring will help mentees to set clear goals and achieve them. For mentors, mentoring will be a reciprocal learning experience that enhances their career growth and learning curve. A successful mentorship requires time, effort, and dedication of both mentor and mentee, which in some cases people cannot afford. To overcome the challenges mentoring has gone beyond the face-to-face traditional meetings which is replaced with online connecting is defined as virtual mentoring.

The term early warning systems (EWS) can be used in different use cases such as in disaster management, security, economics, and finances. However, the use of Educational Early Warning Systems (EEWS) has also now become a prominent use case. Early warning systems in education can be defined as a system which is based on student data that can be used to identify students who exhibit lesser behavior or academic performances that puts them under the category of at-risk students in the education system ("Issue Brief: Early Warning Systems," 2016).

According to Slavin and Madden (1989), the term "at-risk student" can be defined as "someone who faces the risk of not obtaining the necessary level of education to finish university and low achievement, grade retention, behavioral issues, low attendance, low socioeconomic status, and attendance at universities with a high proportion of impoverished learners are risk factors".

With a correct set of indicators, EEWS can be beneficial for students to assess their status of working and will be aware of the improvements to be made. Earliest identification of at-risk students in universities will be highly supportive to reduce the number of dropouts and support students to build up their skills and confidence.

EEWS and virtual mentoring are two phases in a student education process, with EEWS identifying students needing support and guidance, and virtual mentoring guiding students towards desired targets. EEWS systems are mostly built up with predictive algorithms (Liz-Domínguez et al., 2019). EEWS uses Learning Management System (LMS) data, student background details, performance indicators, images, engagement indicators, academic results, and Grade Point Average (GPA), among others, as input for the predictions (Liz-Domínguez et al., 2019). Most of the existing EEWSs collect data, predict the at-risk students with the support of techniques such as Machine Learning (ML) or predictive algorithms. Then finally the relevant results will be informed to the stakeholders including students, educators, university management and parents.

In the best practices, it will be obvious that just identifying the risk level early and informing will be sufficient. The necessary steps should be taken or suggested for students or educators to overcome the risk level. Virtual mentoring will be one such ideal solution to suggest students and educators to uplift the condition of students. With virtual mentoring, it would be efficient if we could suggest the best matched mentor and mentee pair. Still there is the requirement to create a single system which bonds both the aspects of identifying at-risk students and to direct them for the correct mentor. Hence this study addresses the need to create frameworks to build a system to identify at-risk students and warn them early and suggest them with a best matched mentor to address the above stated requirement. In the existing systems, identifying at-risk students and finding the ideal match of mentor-mentee pair will be done as two different processes in separate applications and frameworks.

A system merging both EEWS and mentoring frameworks could enhance university quality of education by identifying at-risk students and matching them with the most suitable mentors. However, existing research on merging these frameworks is limited, with only a few studies providing a framework for EEWS and suggesting the best match mentor. A comprehensive system that integrates these two frameworks is highly required, filling a research gap.

This study aims to address the lack of research on having a single framework for EEWS and mentoring by creating an integrated framework to identify at-risk students in university education and suggest the best matched mentor. The study also explores the research question, "to what extent a Machine Learning and knowledge-based framework can be applied in identifying at-risk undergraduate students and suggest a personalized best matched mentor for them?"

Hence, the aim of the study was to develop and validate a conceptual application by creating an integrated framework for the application to identify at-risk undergraduate students and match mentee with best matched mentor. While achieving that, the study has focused on two main objectives as,

1. to identify the requirements for a system to cater to identify at-risk undergraduate students and matching mentors for them
2. to propose an integrated framework with machine learning based at-risk undergraduate identification and knowledge-based mentor matching.

The rest of the paper has elaborated how the above research question, aim and objectives were achieved. The organization of the paper is as follows. In Section 1, reviews different techniques that are closely related to the study. In Section 2, the methodology and the proposed conceptual framework is described in detail. Section 3 discusses the results of the experiment through conducted case studies. Section 4 discusses the findings, and the last section concludes the work while stating the future works. Finally, references for the citations are provided.

Literature Review

When considering EEWS and mentor matching there are ample studies done related to the field. Literature typically about an integrated framework for EEWS and mentor matching was obscure, although studies related to EEWS followed with several technologies and work related to mentor matching were found among related work. Hence, this section reviews some of the studies that have been proposed over the recent years and among them which have been considered as mentors when commencing this piece of work.

Studies in this section can categorized under the classes given below when reviewing with a focus to examine the lacuna.

1. Impact of EEWS in higher education
2. Impact of Mentoring in higher education
3. Integration between EEWS and mentoring

Impact of EEWS in Higher Education

Plak and colleagues (2021) evaluated the suitability of Early Warning Systems (EWS) in higher education on locating at-risk students and minimizing dropout. In a randomized field trial, it was discovered that EWS correctly predicted at-risk pupils but did not lower dropout rates or boost academic achievement. Hence the work points to the need for further feedback and counseling techniques along with a proper EWS. A field study engaging 1,577 students from three faculties was conducted in 2016–2017 at VU Amsterdam in the Netherlands. After passing a digital Dutch proficiency exam authors have asked students to take part in an experiment using EWS-assisted counseling. EWS provided counseling assistance to the intervention group during the first academic year, while the control group received standard therapy. 16.5% of participants completed a follow-up survey, and 15 out of the 31 participating students for counseling sessions were questioned about how they used the EWS dashboard by authors. As stated by authors, machine learning models outperformed heuristic or theory-based estimation models, addressing over-identification issues. Hence, the study has used the gaining momentum of machine learning in predictive modeling, to support teacher tenure decisions and college dropout prediction. The models have been assessed at the beginning of the school year, at the conclusion of six successive study terms, and at the conclusion of the summer break.

A Dutch counseling program in higher education used an EWS to predict student-specific dropout risks. A dashboard showed the system’s predictions, encouraging one-on-one consultations and coaching interventions. The trial, involving 12 bachelor programs at Vrije Universiteit Amsterdam, found no significant impact on dropout rates or credits earned. However, the study’s low response rate and lack of external validity were drawbacks.

Aguilar and colleagues (2014) investigated how academic advisors, who were unfamiliar with the use of data-driven learning analytics tools use them through an Early Warning System (EWS) powered by learning analytics. The findings demonstrate that, despite the EWS's intended use as a tool to help advisors prepare for meetings and spot students who could be having academic difficulties, advisors mostly used it when they met with students. According to the authors, this has added an unexpected audience and had important design implications for advancements in learning analytics. The main goal of the study was to create and implement Student Explorer across different contexts. It was a joint effort between researchers and academic advisers. The Summer Bridge Program was founded in 1975, provides hundreds of students with intensive academic preparation, customized mentoring, and a living environment that fosters community and was used for the setting of the study. Nine academic advisors, five female and four males, served 219 Bridge students, serving 2,500 students during fall and winter terms. They have Masters and Doctoral degrees in various fields. Students were from three courses: a mathematics course for intermediate algebra or mathematical reasoning, an English or Writing course, and a first-year introduction to social science seminar. Data on advisors' use and attitudes towards Student Explorer were collected from user log data, calendar application log data, and a count of 32 pre- and post-Bridge surveys. The study aimed to understand their background, perceptions of student academic orientation, and overall perceptions of Student Explorer's functionality. The study used Student Explorer to track student progress, LMS site visits, individual assignments, and gradebook comments. Bridge advisers participated in a one-hour training session to enhance accuracy and provide necessary adjustments before attempting the online study.

Through the pre-bridge survey it was identified that advisors anticipated discussing most about mathematics in meetings and English and orientation course next in the order. Also, it has been found that the average count of advisors highlights Student Explorer as one of the valuable tools regardless of their usage. Then authors collected usage data generated by Student Explorer related to each advisor for seven weeks. Authors have also found from the study, student interaction with meetings have gradually decreased, however engagement with Student Explorer has increased during the weeks closer to mathematics midterm exams. The study has observed advisors interact with the Student Explorer by using page access parameters and it has been stated that with a grand total of 3035 accesses advisors have interacted and have shown more interest towards students' status. Further, advisors have used the EWS during the meeting times and lesser access before the meeting and after the meeting the access rate was recorded even less. Authors have also conducted post-bridge surveys and results have indicated that the English has climbed up in the anticipated subject order for meetings in the pre-bridge surveys by acquiring the position as same as mathematics. Overall, the study has contributed with a EWS called Student Explorer, intended to support advisors with student achievement details, was found to be primarily used during student meetings, suggesting that unintended users may indirectly use learning analytics interventions.

Impact of Mentoring in Higher Education

Lunsford and colleagues' (2017) scholarly work reviews on mentoring in higher education for undergraduates, graduate students, and faculty members. It explores mentoring in educational contexts in the US, Australia, Canada, New Zealand, South Africa, and the UK. The study synthesizes findings from the past ten years to provide evidence on special populations and program types. The required data for the study were obtained from databases related to Academic Search Complete, EBSCO host, Psychology and Behavioral Sciences, SOCI Index and based on several academic and journal writings. According to the study, recent research on mentoring in higher education in English-speaking countries indicates that informal and formal mentoring is prevalent, with formal mentoring being more frequent for undergraduate students and less frequent for faculty members. Mentoring programs focus on underrepresented groups, research, professional and peer mentoring, and early career faculty needs. Further, benefits of mentoring vary depending on the population involved. Also, as emphasized by the authors, future research should explore equivocal results, clarify mentoring relationships, and study effects at under-represented career stages.

Dorner and colleagues (2020) explored online international mentoring for faculty in geographically distant universities. The researchers analyzed interview data from 30 mentees using an inductive analysis technique to understand how online mentoring aids young academics in their growth as inexperienced teachers. The study aims to provide new generations of faculty with strategies to adapt to a specific academic environment and critically examine the limits of teaching and knowledge development influenced by physical location. The results revealed diverse conceptions of the mentoring process, the mentor's function, and the possibility for change in business relationships. A model of transformational experiences was developed to explain the various cycles of professional growth in an online faculty mentoring program. The study uses a grounded theory methodology to analyze qualitative research data and examines the consistency of interviews.

According to the study, mentees have various conceptions about mentoring, the mentor's function, and the transforming power of business connections. It also points up difficulties with faculty mentoring programs conducted online, such as how mentors' sincerity may be harmed by their physical distance. Also, the results from the study are not generalizable and may need to be changed in the future to account for variances in mentoring for mentees who have not had any formal teacher preparation. Further, it was ensured that the study's theoretical framework and program design components are relevant and pertinent for those operating conventional and online faculty mentoring programs abroad by providing these descriptions.

Integration Between EEWS and Mentoring

In the study titled as "Efficacy of the Check & Connect Mentoring Program for At-Risk General Education High School Students" (Heppen et al., 2017), has evaluated the effectiveness of Check & Connect with general education pupils who has exhibited the early warning indicators of high school dropout risk. In the study, authors have taken the student sample based on attendance, behavior, and course performance in Grades 8 and 9, where the study has used a 553 such identified students' population with the lowest estimated likelihood of graduating on time. In the study, students have been randomized to either not receive a Check & Connect mentor for three years, beginning the summer following Grade 9, or to receive one for three years. The result of the study indicates that the program was executed faithfully, except for students who departed from district schools. Further, the study has found that there were no statistically significant effects of Check & Connect on graduation rates, academic advancement rates, dropout rates, or engagement metrics. However, study has emphasized the importance of identifying at-risk students early and directing them for mentoring.

Grewe and Kleiner (2023) offered a model of successful integration of evidence-based mentorship practices during the first year of university education with a program conducted at Utah State University. The program's mentoring component has created to cope with the issue of first-year students leaving early for their second year of study. In the program, every student has received faculty mentoring through the proposed strategy. Further, through the proposed model special attention has been given to the most vulnerable students to support those who do not have the social or educational resources to seek out faculty mentorship on their own. Through the proposed model, evaluation results have shown that it is necessary to keep the program rigorous and prioritize the needs of most vulnerable students, to provide them with the highest quality of high-touch mentoring. Moreover, through the study authors have contributed with an easily adaptable, evidence-based model that can be successfully implemented at any other college.

The study titled as, "Towards Requirements for Intelligent Mentoring Systems" (Kravčík et al., 2019), has addressed the research question "How can we design educational concepts that enable a scalable individual mentoring in the development of competences?" by the authors. When addressing the above gap, the study has aimed to develop knowledge services for an automatic realization of parts of the individualized mentoring process. To achieve the goal, the study has referred to existing systems and analyzed the features and identified requirements for an intelligent mentoring system. According to the findings of the study, to provide metacognitive support, lifelong mentoring, affect detection and accurate predictions, an intelligent mentoring system would be beneficial. Further, the authors have stated that an intelligent system will be required in three phases as in preparation, in the learning process and in the follow-up.

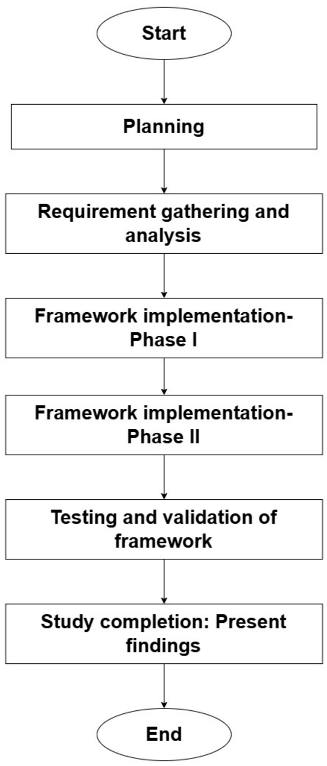
Conclusive Remarks of Literature Review

Recent studies on e-learning and mentoring in EEWS have evaluated the success of mentoring programs, with some discussing the mentor matching process. Qualitative approaches are common, and AI and ML are under investigation for selecting the best mentor. Accurate predictions of at-risk students should consider social, economic, and behavioral factors. The research highlights the need for an integrated framework for merging EEWS and mentoring, with a case study to validate these approaches.

Methodology and Conceptual Framework

The study contributes with an integrated framework named as "GRADGROOM," grooming a graduate, for an application to identify at-risk students and match mentee with best matched mentor. Hence, the methodology of the research has two phases. Phase 1 of the study was to create EEWS to identify at-risk undergraduates. The results obtained with phase 1 will be contributing to initiating phase 2 of the study. Phase 2 was designed to make suggests for the best matched pair of mentor and mentee. The overall methodology of the study has been illustrated in Figure 1 given below.

Figure 1
The Methodology of the Study



Planning

In the planning stage of the study, scope and objectives were decided. Initially a sound literature review was conducted in the selected field to identify an existing gap to address through the study. Once the lacuna was identified, aims and objectives are structured while clarifying the scope and contribution. Further, the rest of stages of the study were designed during the planning stage.

Requirement Gathering and Analysis

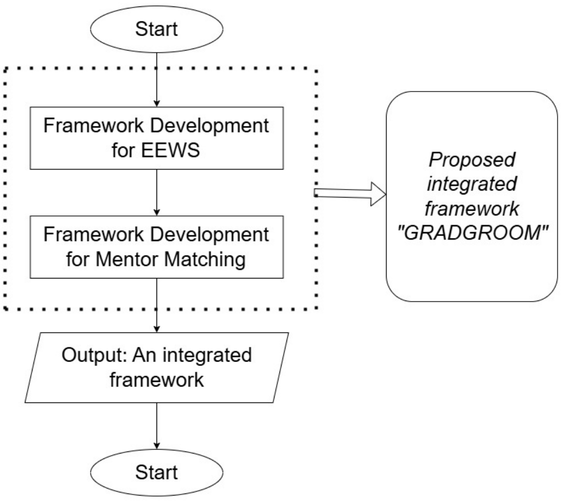
In this stage of the study, the requirement for the specific need was evaluated with existing studies as secondary sources. Further, the relevant stakeholders, undergraduates, lecturers, and university administrations were questioned about the real requirement for an integrated framework. Hence, once the research gap was identified and clarified, the study continued.

Framework Development

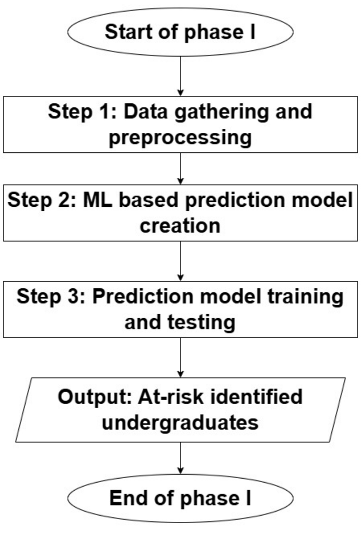
This was the main stage of the study, where the contribution of the research was produced. In the framework development, it was done to address the identified gap in the existing works. The process of framework development contained various steps in two main phases as shown in Figure 2 below.

Figure 2

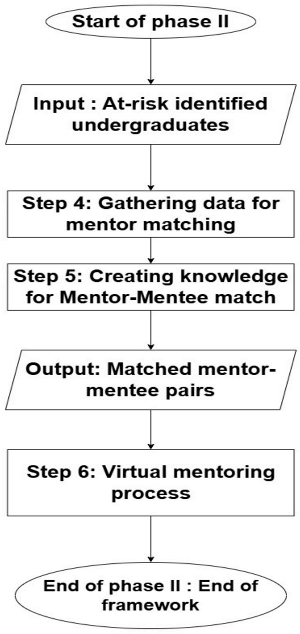
Process of Framework Implementation: (a) Proposed Integration for Frameworks, (b) Phase I Implementation and (c) Phase II Implementation



(a)



(b)



(c)

When the framework was developed, integration of two frameworks for EEWS and mentor matching was the main concern. Hence initially the framework for EEWS was designed and the output generated from the framework was passed to the next design which was a framework for mentor matching as the input. Hence, the study was able to integrate the frameworks to a single framework, where it has been named and introduced as "GRADGROOM." GRADGROOM hence will be supported to identify at-risk students, and then best matched mentors will be suggested for the at-risk students.

According to the proposed methodology, initially the past data will be collected and preprocessed. Then the data will be used to train and test the ML model for EEWS. Hence the at-risk predictions for the students will be done by the trained model and then the identified at-risk students' mentor preferences will be collected. With the collected data from the mentors and mentees, using the knowledge created, the best matched mentor pair will be selected, and the virtual mentoring process will be conducted.

Conceptual Framework

The study has proposed the framework to identify at-risk students in the university systems following undergraduate programs in Technology and Engineering (TE). The undergraduate programs can be categorized into two categories as Science, Technology, Engineering, and Mathematics (STEM) and non-STEM. However, the proposed EEWS framework was designed to be focused for STEM subjects' undergraduate programs since it has identified with much research that STEM has major challenges to face with its low student enrollment and high attrition rates and due to these challenges more students tend to drop the program or switch to non-STEM programs (Sithole et al., 2017). Further it has also been observed that a study conducted over with 110,000 student records has shown clearly that getting a grade of C or lower in an introductory STEM class has a been heavily impacted on the likelihood of underrepresented college students earning a STEM degree (Sholtis,2022). Like this evidence, many other works also have emphasized the high rate of dropouts and challenges in STEM undergraduate programs and critically analyzed the requirement of an EEWS for STEM programs (Bernacki et al., 2020; Yu & Wu, 2021).

Hence, in the study, it has chosen STEM undergraduate programs as its scope and based on the practical considerations it was limited to TE spectrum of STEM. Further, studies have also shown that it would be appropriate to predict at-risk students earlier in degree programs to prevent the dropping out of students (Adnan et al., 2021; Arvind JNR Kumar, 2018). Based on the facts, the proposed framework for EEWS is used to predict at-risk students from first year and second year courses.

With the above-mentioned limitation, the proposed framework "GRADGROOM," can be applied for any TE stream programs in universities. The proposed framework consists of 6 steps, each step completion will provide the university an effective procedure to control the drop-out rates. In the step 1, university should gather the data records of the past graduated and drop-out students. This data can be collected based on 28 parameters suggested by the proposed framework as elaborated in the below paragraphs where the step-by-step procedure has been explained. Thus, collected data can be preprocessed and cleansed to use as test and train data for the ML model. In step 2, the study has proposed to use KNN, RF, and SVM algorithms to ensemble through stacking with the Meta model based on DT for the at-risk identification. By step 3, the model can be trained and tested from the data collected from step 1, to prepare the ML model to be applied for the at-risk prediction of the current students. After this step, the university can collect data for the same 28 parameters again from the current set of students who needed to be classified as at-risk and well-performing. Then, this dataset can be fed to the ML model to acquire the prediction results.

After obtaining the results from the prepared ML model from step 3, universities then can focus to uplift the status of at-risk identified students. As to take immediate actions, the framework continues to the step 4 where the mentor-mentee matching framework has been merged with the at-risk identification framework. In step 4, focusing on the at-risk identified undergraduate, university can collect few more parameters to use along with the data of the universities' mentee pool to suggest the best matched mentor to them. Step 5 will be using the data collected in step 4 to match the best matched mentor-mentee pair. This step utilizes the knowledge-based logic program and will provide the best matched pairs to the university, facilitating the mentoring process. Based on the suggestions, the university hence can conduct the last step of the proposed framework where a virtual mentorship program will be conducted.

Most importantly, the framework can be reused from step 1 if the university needs to have more accuracy for the predictions with latest passed out student details for every upcoming batch's at-risk prediction. Whereas can be reused only from step 3, for upcoming student batches if the university is satisfied with the current accuracy of the prediction. Also, if students are not happy with the suggested mentor through the framework, step 5 can be repeated or can use the next best matched suggestion. This flexibility of the proposed framework can be highlighted as one of the biggest strengths of it. The section given below, elaborate each step with technical details.

Step 1: Data Gathering and Preprocessing

As the first step, past data gathering from the TE stream University undergraduate programs should be done to create a dataset for the at-risk prediction. The collected data must serve as test and train data for the ML based prediction model. Framework has intended to collect data TE subject(s) in the first year and the second year of the degree program to predict at-risk students for the course. Through literature it was observed that LMS data plays a prominent role in predicting at-risk students (Arizmendi et al., 2022; Howard et al., 2018; Macfadyen & Dawson, 2010; Osborne & Lang, 2023). Further, it has also identified that students' assessment scores, engagement intensity and time dependent factors are impacting more in the online learning environments (Marwaha & Singla, 2019). Moreover, studies have emphasized that socio-economic status also matters for students' education (Slavin & Madden, 1989).

Su and colleagues (2022) conducted a study to identify factors to be used in at-risk student prediction with ML techniques. Study has stated that academic factors should be considered as they define academic performance such as marks. Also, authors have stated demographic data and social and behavioral data should be equally considered as with demographical data. Because with demographic data, it can be used to define the characteristics of a person such as race, age, and income while social and behavioral can be used to define the lifestyle and habits of a person such as religious believes, family income, and so on.

Hence in this study, the framework has been designed to take 28 parameters covering academic, LMS, demographics and behavioral features. The details of the parameters are described in Tables 1, 2, 3 and 4 given below.

Table 1
Academic Parameters for At-Risk Prediction in EEWS

Category	*Parameters*
Academic Features	End exam grades and assessment grades
	Subject preferences
	Lecture clarity
	Lecture attendance

Table 2
Academic Parameters for At-Risk Prediction in EEWS

Category	*Parameters*
Demographic Features	Gender Age District Secondary education English literacy • O/L English results • A/L English results Advanced level results Family background • Father's occupation • Mother's occupation • No of siblings • Family income status Relationship status Monthly living expenses as a student Job status • Full time student • Part time worker • Full time worker (Part time studying) Special conditions • Medical issues • Economic issues • Family dedications • None Current stay: Hostel/ Home/ Boarding

Table 3
LMS Parameters for At-Risk Prediction in EEWS

Category	*Parameters*
Features from LMS	LMS active time No of on-time/delayed submissions Forum activeness • No of forum posts/replies • No of words in forum posts/replies No of account logins for the semester

Table 4
Behavioral Characteristics for At-Risk Prediction in EEWS

Category	*Parameters*
Behavioral characteristics	Monthly Internet usage Daily mobile screening time Daily study time Mood controlling skills Reputation at university- Character and conduct Extra-curricular activities Stress level handling and controlling

Once these parameters are collected, then the data are preprocessed and purified to form training and testing data for the model.

S***tep 2: ML Based Prediction Model Creation***

During the literature survey, it was identified a range of algorithms that can be used for the at-risk student prediction purpose such as Support vector machine (SVM), decision trees (DT), random forest (RF), Naive Bayes classification, K- nearest neighbors and Neural networks (NN) (Arizmendi et al., 2022; Baneres et al., 2019; Howard et al., 2018; Plak et al., 2021). Several existing studies have shown NN, RF and DT algorithms give high accuracies when predicting at-risk students (Akçapınar et al., 2019; Marwaha & Singla, 2019). Further, KNN has also been found as highly accurate for the purpose (Xu et al., 2023). SVM was also highlighted through studies as one of the best algorithms which work well with a small dataset (“Matching Mentors with Mentees”, 2022).

Hence, this framework is intended to apply KNN, RF, and SVM algorithms to ensemble through stacking with the Meta model based on DT for the framework to predict at-risk students.

Step 3: Prediction Model Training and Testing

The prediction model was trained with 80% of the data and tested with 20% of the gathered data. With these steps, the first phase of the study will be completed by designing the framework for the EEWS.

In the second phase, the above predicted at risk students' data will be used in the mentor-mentee matching program. Hence the two frameworks will be integrated at this phase. Framework development for mentor-mentee matching will contain below steps. When finding the best matched mentor and mentee pair, existing works have shown the importance of considering deep-level characteristics (personal traits), surface similarities and individual experience (Keramidas et al., 2022). Hence in the proposed framework the three categories of features mentioned above, and Learning Goal Orientation (LGO) were considered.

Step 4: Gathering Data for Mentor Matching

In this step framework is designed to gather more data from at-risk predicted students and from mentors. The gathered data will be used to match the mentor-mentee pair for the virtual mentoring process.

To obtain the deep-level personal traits, Five-Factor Model (FFM) also referred to as "Big Five," which is one of the widely accepted frameworks in psychology, was used. The selected framework uses five factors acronymic as OCEAN: Openness to experience, Conscientiousness, Extraversion, Agreeableness and Neuroticism to understand the personal traits. The framework was selected as it covers a broad dimension to describe human personality (Garcia C., 2019; Zhao et al., 2021).

For surface similarities the framework has considered gender and ethnicity and for individual experience the framework has considered education qualifications and career experience. Further, Learning Goal Orientation (LGO) was used to understand the students' underlying attitude and approach toward learning. It was selected as one of the factors in mentor matching because it encompasses how individuals perceive and engage with learning activities, their motivations, and their strategies for learning and skill development (Lechuga & Doroudi, 2022).

Hence in this study, the framework has been designed to parameters covering deep-level characteristics, surface similarities, individual experience and LGO features. Deep-level characteristics and surface similarities were considered for both mentors and mentees while individual experiences were taken from mentors and LGO was considered from mentees' perspective. The details of the parameters are described in the Tables 5, 6 and 7 given below.

Table 5
Surface Level Parameters for Mentor-Mentee Matching

Category	*Parameters*
Surface level similarities	Gender
	Age
	Ethnicity
	Religion

Table 6
Interest Identification Parameters for Mentor-Mentee Matching

Category	*Parameters*
Individual experience	Current job title
(From mentors)	Years of experience
	Previous mentoring experience
Learning Goal Orientation (LGO)	Focus on acquiring knowledge
(From mentees)	End goal of learning
	Motivations for learning

Table 7
Deep Level Parameters for Mentor-Mentee Matching

Category	*Parameters*
Deep level characteristics	Openness to experience Example: • Open-mindedness • Imaginative skills • Curiosity level • Creativity level Conscientiousness Example: • Level of organization of work • Reliability of work • Goal-oriented Extraversion Example: • Outgoingness • Energy of working • Sociable • Assertive Agreeableness Example: • Empathy • Kindness • Cooperation Neuroticism Example: • Anxiety • Depression • Anger • Mood swings • Stress handling

Surveys were conducted with questionnaires made based on the above factors which was verified from an educational counselor to gather the relevant data for mentor-mentee matching.

Step 5: Creating Knowledge for Mentor- Mentee Match

With the gathered data through pre surveys knowledge-based mentor-mentee matching was proposed in the framework. When paring two people together several techniques such as complementary paring, similarity paring, task-specific paring and contextual paring can be made. In educational systems mixed approaches will give enhanced results as observed through existing works (Pursell, 2024). Further, it was identified that among various other techniques to match mentees with mentors, skill-based compatibility matching with similarities will be ideal for the proposed framework (Pursell, 2024). Then from the collected designed to use knowledge-based logics to match the best suitable mentor and mentee pair. All the collected parameter categories from mentors and mentees were given score values and based on the scores the feature categories were labeled as low, medium, and high with the calculation of tertile. Hence, logics were made based on skill-based compatibility matching with similarities. Then with an educational counselor's consultation the logic was verified. With the knowledge, matched pairs will be produced at the end of the framework.

Step 6: Virtual Mentoring Program

Since the mentors and mentees cannot be limited with the geographical locations and other constraints, the framework suggests a virtual mentoring program which will be conducted throughout the year. Further, the framework suggests evaluating the growth of students frequently after few mentoring sessions according to the university requirements and through the feedback obtained from mentors and mentees, if required the framework can be reused to rematch the mentors and mentees with the changed or updated mentee requirements.

Testing and Validation of the Framework: A Case Study

To validate the above proposed integrated framework through the study, two case studies were conducted as explained in detail in the Section 3. Real data from a local higher education institute in Sri Lanka was collected for two different subjects in the first year and the second year of two programs of degrees and proposed conceptual framework testing and validation. Hence by collecting post survey results from the students the proposed integrated framework was validated.

Present Findings: Results Analysis

With the case studies and obtained results, the effectiveness of the selected algorithms for EEWS were compared individually and after stacking them to a single model. Further, through the combined model at-risk student prediction results were produced and then the mentor-mentee pair matching was conducted. To produce results to evaluate the effectiveness of integrating the at-risk prediction and effective and efficient mentor-mentee matching, selected samples of at-risk predicted and non-at-risk predicted students were engaged in pre-survey and a post-survey during the virtual mentoring period. Hence the obtained results were compared to produce results to show the effectiveness of at-risk prediction and navigating them for a proper mentoring program through an effective mentor matching. Findings of the study has been elaborated in Section 4.

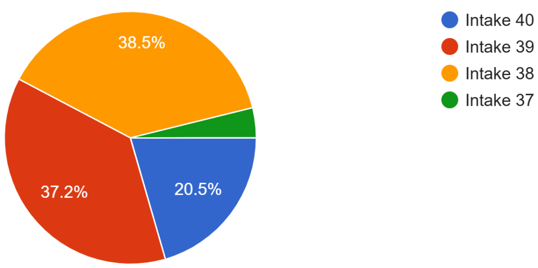
Testing and Validation of Framework- Case Studies

To evaluate the proposed framework two case studies were conducted and test results were produced. Case studies were conducted with the data gathered from one of the local higher education institutes in Sri Lanka and all the data used were anonymized during the writing and publications to preserve the ethical considerations of university and students. For the case study computer science degree stream and software engineering degree stream students of 4 intakes were selected as the sample. The sample was selected considering the TE subject streams in the Faculty of Computing of the higher education institute.

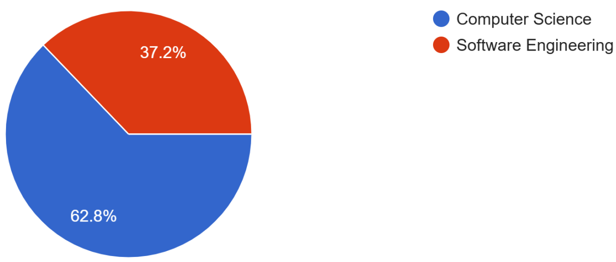
Demographic data from all the four years were taken and in addition to that subject specific data from relevant student groups were taken through questionnaires as shown in a few samples in Figure 3 to Figure 5 given below. The questionnaire consists of 31 questions, and they were finally merged with the subject specific data gathered during the case studies to form the final 28 parameters for the prediction.

Figure 3

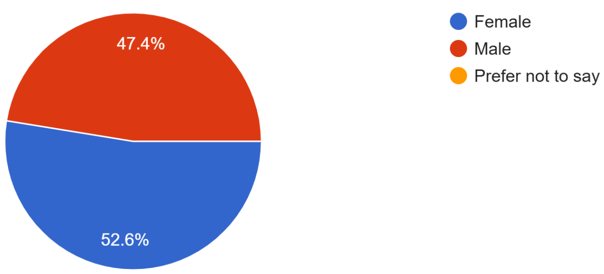
Basic Data Gathered About Students: (a) Academic Intake, (b) Subject Stream and (c) Gender Distribution



(a) Academic intake



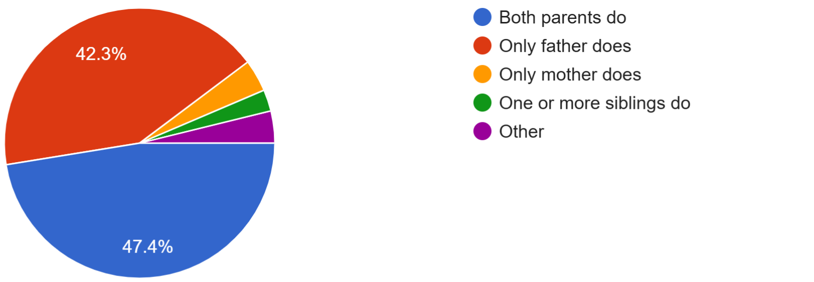
(b) Subject stream



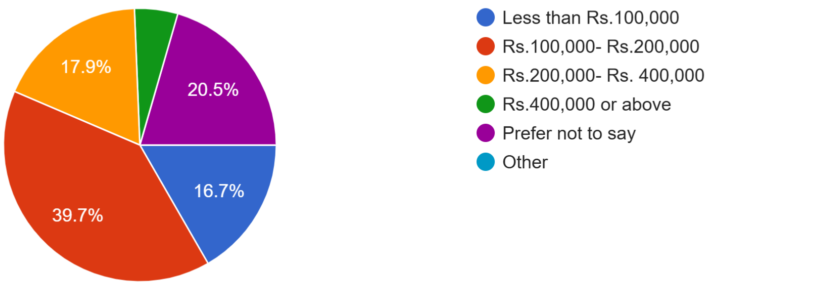
(c) Gender distribution

Figure 4

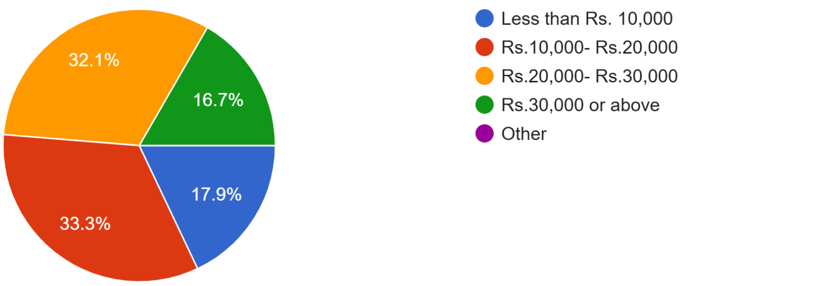
Economic Related Factors to the Student: (a) Economic Providing for the Family, (b) Monthly Family Income, (c) Monthly Expenses of the Student and (d) Economical Provider for the Student



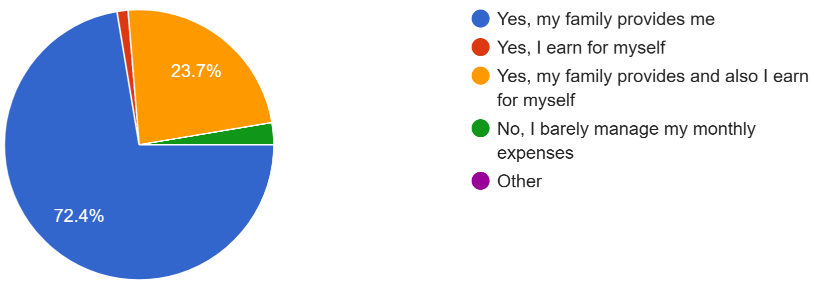
(a) Economic providing for the family



(b) Monthly family income



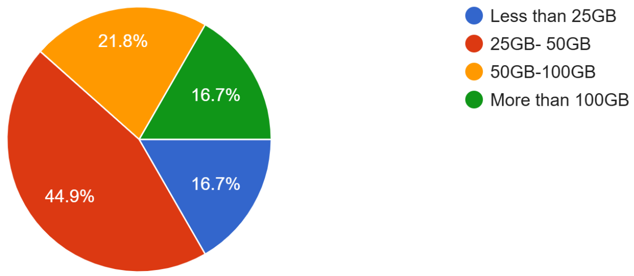
(c) Monthly expenses of the student



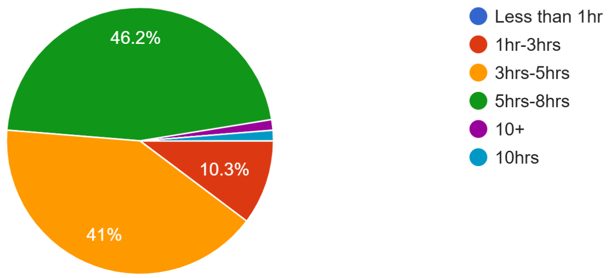
(d) Economical provider for the student

Figure 5

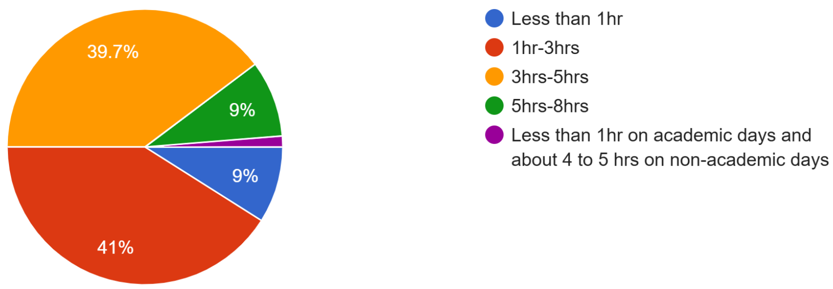
Details about Study and Internet Usage: (a) Monthly Internet Usage, (b) Daily laptop/mobile device screen time and (c) Daily study time



(a) Monthly Internet usage



(b) Daily laptop/mobile device screen time and



(c) Daily study time

Moreover, for the second phase of the study for mentor-mentee matching, five mentors were selected from academia and industry randomly covering the fields of software engineering as the selected subjects were both relevant to the field of software engineering. This chapter will be further elaborating the details of the case studies conducted.

Case Study 1: Predicting At-Risk Undergraduate Students at First Year

In the case study 1, one the 1st year undergraduate subjects, Fundamentals of Programming (FP) was selected. Other than the collected demographic data from students, subject specific academic data such as grades and LMS data for the specific selected subject from all four years were collected. The collected data of the 1st year subject of current 2nd, 3rd and 4th year students have been used as testing and training data for the ML model.

That data then has been split into testing and training dataset and actual 1st year students' data was used as the validation dataset. Collected data test and train have used as the 80% training data and 20% testing data to train and test the ML model for at-risk prediction. The data collected from first year students were taken as a validation dataset and fed to the model to get the results and compared with the already known labels. Both train-test dataset and validation dataset had 28 columns excluding the label and serial number. Train-test dataset for the case study two had 121 data rows and the validation dataset had 36 rows.

With the at-risk predicted students from the framework, a sample of five students were taken randomly to conduct the mentor-mentee matching. Initially additionally required data for mentor-mentee matching was collected through a questionnaire. Questionnaire contained 60 questions covering surface level personal traits, deep level personal traits and LGO data. Then the data gathered from mentees were pre-processed while giving scores for each category of features out of 100%. From the selected five mentors, data were collected through another questionnaire where it contained 60 questions covering surface level personal traits, deep level personal traits and professional experiences and score values were given. Both questionnaires were made based on FFM-OCEAN to gather deep level personal traits and questions were verified through an educational counsellor.

Data gathered from mentors and mentees were then labelled as low, medium and high with the calculation of tertile of the score values. With the score values for deep level characteristics and surface level characteristics mentor-mentee pairs were matched according to the similarly. Experience level from mentors were matched with the LGO of mentees in a way to uplift the student improvements. High experienced level mentees were matched with high and low levels of mentees with LGO. Low and medium experienced mentors were matched with medium LGO mentees.

Finally, according to the features, matched pairs of mentors and mentees were matched from the framework and were given opportunities to conduct three mentoring sessions each of not less than 30 minutes weekly. Then post feedback gathering interviews were conducted from the mentees who were paired and to compare the improvement at-risk identified non-mentored students were also questioned. Finally, the feedback from two parties were analyzed to produce results to evaluate the success of the proposed integrated framework, which extends the process of at-risk prediction up to mentoring.

Case Study 2: Predicting At-Risk Undergraduate Students at Second Year

In case study 2, the above same procedure was conducted for one of the 2nd year subjects, Data Structure and Algorithms (DSA). After collecting the subject specific academic data and LMS data from 2nd, 3rd and 4th year students the dataset was merged with demographic data. Then the current 3rd year and 4th year students' data for the selected 2nd year subject were split into training and testing in percentages 80% and 20% to train and test the ML model for at-risk prediction. Then the data collected from current 2nd year students were taken as a validation dataset and fed to the model to get the results and compared with the already known labels. Both train-test dataset and validation dataset had 28 columns excluding the label and serial number. Test-train dataset for the case study two had 66 data rows and validation dataset had 55 rows.

With the at-risk predicted students from the framework, a sample of five students were taken randomly and mentor-mentee matching was conducted. Initially additional data for mentor-mentee matching was collected through a questionnaire and pre-processed while giving scores as in the previous case study, “case study 1”. While using the same knowledge created in the above case study, mentors were matched with mentees.

Finally, the paired-up mentors and mentees from the framework were given the opportunity to conduct three mentoring sessions each of not less than 30 minutes weekly. Then post feedback gathering surveys were conducted to evaluate the success of the proposed integrated framework.

Results and Discussion

After executing the framework and predicting at-risk undergraduates and matching best mentor to them, mentoring sessions for both cases were conducted.

The sessions were conducted virtually in adhering to video calls to make the mentors and mentees closer to each other in despite of the location barriers. Five selected mentors participated in the process and among the randomly selected five at-risk predicted students from each sample were considered in matching. In the matching, based on the knowledge created each mentor was matched with a mentee resulting in five out of ten intended mentees to be matched for mentoring. Finally, the feedback was gathered from five mentees who participated in the mentoring sessions and other five remaining at-risk predicted students who were not mentored. Gathered feedback was analysed to understand the impact of integrating the two frameworks.

At-risk Undergraduate Student Mentoring and Not Mentoring Impact

Individual feedback received through interviews based on 5-6 main questions. Through the interviews mentees highlighted that the mentoring program was highly impactful, and they hope it would have continued during your university period. Student 15 and 32, highlighted that they were motivated with the sessions and the feedback of mentors and not as they do their studies alone, they felt guided with short term goals. Given below are the summaries of the feedback of the mentored students.

Summary of individual student feedback

Student 24: Student has rated the program 8/10. Also stated that initially the communication was not easy but within a few minutes it was comfortable, and the mentor was friendly even more than the student had thought to be. Further, the mentor has supported in creating study goals and student had achieved them to a satisfactory level as the mentor has continuously checked the progress. Overall, students highlighted the importance of the sessions and were willing to have continuous sessions thereafter too.

Student 15: Student was highly satisfied about the input from the mentor and rated the program 9/10. Further, mentioned that goals were set with the support of mentor and mentor motivated student which the student felt secured and guided. Student agreed with most of the feedback given by the mentor, however student stated the mentoring period was not long enough to experience it well.

Student 32: Student rated the program 7/10 and stated the mentor was good but could have been more specific when giving instructions. Further, the student stated that he was satisfied with the feedback and the student also stated that the program was not long enough. It could have been more effective if it was held for a long period and even during this short period student was able to gain a lot according to the student's perspective.

Student 4: Student rated 9/10 for the program and stated that identifying as at-risk was always a worry for the student. However, through the given guidance student got motivated to work well in future. Student said that she was able to clear out many doubts of her which she could not solve during her classroom with a close attention. Student also stated as the others that the program could have been longer to get the best impact.

Students appreciated mentors' support in setting goals and motivating them. They enthusiastically worked on tasks for the week, preventing them from missing work if they had only set goals and worked for them.

The study reveals that non-mentored students often struggle with receiving close attention during lectures, leading to doubts about studying, exam preparation, and industry. Some students are at-risk and focus more on studies, while others are not satisfied with their progress. The study also found that students lack the proper idea of setting academic and life goals without a mentor's support. However, two students were not worried about being identified as at-risk and were comfortable working slowly with their normal routines, even if they lost the mentoring opportunity. The feedback reflects the perspectives of non-mentored students and highlights the importance of mentorship in addressing these challenges.

When comparing the proposed framework with the existing frameworks, the "Check and Connect" program presented by (Heppen et al., 2017), shows limitations in the study by highlighting the importance of early identification of at-risk students and importance of collecting the correct set of data for prediction. And this framework was for school education.

Hence in the newly proposed system, it has been suggested to run the new framework in early of the university academic programs preferably in the first and second academic years. In the study (Grewe & Kleiner, 2023), importance of mentoring has been highlighted, however the framework has not prioritized the at-risk students. It has made the administrative and procedure for mentoring more complex. Hence, in the proposed new framework from this study has focused on at-risk identified students making the procedure effective and efficient. Further, compared to framework proposed in the study (Kravčík et al., 2019), the flexibility of the newly proposed framework can be stated as a prominent difference. Further, the many existing studies only having individual frameworks either for mentoring or EEWS. Hence, the "GRADGOOM" framework has attempted to fulfill the existing research gap.

The study concluded that extending at-risk predicting frameworks, EEWS, with a mentoring process is crucial for at-risk undergraduate students. A proper mentoring process can motivate them to return to the correct study route, and delayed mentoring can fade their enthusiasm. The study also highlighted the varied study routines and the impact of mentoring on students. However, the study had limitations, such as the need for a second case study to compare results, more data collection to avoid overfitting and error proneness, and the inclusion of all university batches. The mentor pool could have been expanded to include all at-risk identified students, and mentoring duration could have been extended due to time constraints. By addressing these issues, the study could have been more effective. Overall, the study highlights the importance of implementing mentoring processes to improve student outcomes.

Conclusion

While achieving all the objectives and yielding many primary findings, the study has correctly proven its aim which was to evaluate the success of the proposed integrated framework for EEWS and mentoring. Overall, through the case studies and feedback gatherings, study has emphasized the importance of extending the at-risk predicting frameworks, EEWS with a mentoring process. Since a proper mentoring process is conducted immediately after being identified as at-risk undergraduate students, students will be taking initiative to reshape their study patterns to align with the correct route of studying. With the extension of the framework of EEWS up to mentoring and creating the integrated framework as "GRADGROOM" with the proposed methodology, the above stated requirement can be clearly addressed.

The proposed solution has covered practical aspects related to the identified problem and has been able to solve the problem up to a considerable level as proven through the case studies. Hence it can be stated that with a proper set up for the designed experimental setup this framework can be used at any higher educational institute to address their specific requirements. With the findings, it can be concluded that the early prediction of at-risk undergraduates and directing them for a proper mentoring program would be ideal and have emphasized the importance of the integration of the two aspects.

References

Adnan, M., Habib, A., Ashraf, J., Mussadiq, S., Raza, A. A., Abid, M., Bashir, M., & Khan, S. U. (2021). Predicting at-Risk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models. *IEEE Access*, 9, 7519–7539. https://doi.org/10.1109/ACCESS.2021.3049446

Aguilar, S., Lonn, S., & Teasley, S. D. (2014). Perceptions and use of an early warning system during a higher education transition program. https://doi.org/10.1145/2567574.2567625

Akçapınar, G., Altun, A., & Aşkar, P. (2019). Using learning analytics to develop early-warning system for at-risk students. *International Journal of Educational Technology in Higher Education*, *16*(1). https://doi.org/10.1186/s41239-019-0172-z

Arizmendi, C. J., Bernacki, M. L., Raković, M., Plumley, R. D., Urban, C. J., Panter, A. T., Greene, J. A., & Gates, K. M. (2022). Predicting student outcomes using digital logs of learning behaviors: Review, current standards, and suggestions for future work. *Behavior Research Methods*, *55*(6), 3026–3054. https://doi.org/10.3758/s13428-022-01939-9

Arvind JNR Kumar, R. (2018). Early Warning System in Education. *International Journal of Advanced Research and Publications*, *2*(9), 115–116.

Baneres, D., Rodriguez, M. E., & Serra, M. (2019). An Early Feedback Prediction System for Learners At-Risk Within a First-Year Higher Education Course. *IEEE Transactions on Learning Technologies*, *12*(2), 249–263. https://doi.org/10.1109/TLT.2019.2912167

Bernacki, M. L., Chavez, M. M., & Uesbeck, P. M. (2020). Predicting achievement and providing support before STEM majors begin to fail. Computers and Education/Computers & Education, 158, 103999. https://doi.org/10.1016/j.compedu.2020.103999

Dorner, H., Misic, G., & Rymarenko, M. (2020). Online mentoring for academic practice: strategies, implications, and innovations. *Annals of the New York Academy of Sciences*, *1483*(1), 98–111. https://doi.org/10.1111/nyas.14301

Garcia C. (2019). The Moderating Role of Goal Orientation in Gamified Instruction: An Extension of the Theory of Gamified Learning. Master of Science in Psychology Dissertation, Illinois Institute of Technology, ProQuest Dissertations Publishing.

Grewe, J., & Kleiner, H. (2023). Chapter 19- The Connections Program: Integrating Mentoring into the First-Year Experience. In Making Connections. Retrieved from https://digitalcommons.usu.edu/makingconnections/21

Heppen, J. B., Zeiser, K., Holtzman, D. J., O'Cummings, M., Christenson, S., & Pohl, A. (2017). Efficacy of the Check & Connect Mentoring Program for At-Risk General Education High School Students. *Journal of Research on Educational Effectiveness*, *11*(1), 56–82. https://doi.org/10.1080/19345747.2017.1318990

Howard, E., Meehan, M., & Parnell, A. (2018). Contrasting prediction methods for early warning systems at undergraduate level. *The Internet and Higher Education*, 37, 66–75. https://doi.org/10.1016/j.iheduc.2018.02.001

Issue Brief: Early Warning Systems. (2016). In National Survey on High School Strategies Designed to Help At-risk Undergraduate Students Graduate (HSS) Project. U.S. Department of Education, Office of Planning, Evaluation and Policy Development Policy and Program Studies Service. Retrieved from https://www2.ed.gov/rschstat/eval/high-school/early-warning-systems-brief.pdf

Keramidas, N. L., Queener, J. E., & Hartung, P. J. (2022). Forming mentoring relationships in graduate education: The role of personality. *Australian Journal of Career Development*, *31*(2), 118–129. https://doi.org/10.1177/10384162221107972

Kravčík, M., Schmid, K., & Igel, C. (2019). Towards Requirements for Intelligent Mentoring Systems. https://doi.org/10.1145/3345002.3349290

Lechuga, C. G., & Doroudi, S. (2022). Three Algorithms for Grouping Students: A Bridge Between Personalized Tutoring System Data and Classroom Pedagogy. International *Journal of Artificial Intelligence in Education*, *33*(4), 843–884. https://doi.org/10.1007/s40593-022-00309-y

Liz-Domínguez, M., Caeiro-Rodríguez, M., Llamas-Nistal, M., & Mikic-Fonte, F. A. (2019). Systematic Literature Review of Predictive Analysis Tools in Higher Education. *Applied Sciences*, *9*(24), 5569. https://doi.org/10.3390/app9245569

Lunsford, L. G., Crisp, G., Dolan, E. L., & Wuetherick, B. (2017). Mentoring in higher education. CU FIND. https://doi.org/10.4135/9781526402011.n20

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers and Education/Computers & Education*, *54*(2), 588–599. https://doi.org/10.1016/j.compedu.2009.09.008

Martha. (2022, February 15). Role and Importance of Mentoring in Education. Sanjeev Datta Personality School. Retrieved from https://sanjeevdatta.com/importance-of-mentoring-in-education/

Marwaha, A., & Singla, A. (2019). A Study of Factors to Predict At-Risk Students Based on Machine Learning Techniques. In Advances in intelligent systems and computing (pp. 133–141). https://doi.org/10.1007/978-981-13-8618-3_15

Matching mentors with mentees. (2022). Human Resource Management International Digest, *30*(7), 43–45. https://doi.org/10.1108/HRMID-08-2022-0231

O'Connell, R. (2024, April 9). Benefits of Mentoring Programs and Mentorship | Guider AI. Guider AI. Retrieved from https://guider-ai.com/blog/mentoring-benefits/

Osborne, J. B., & Lang, A. S. (2023). Predictive Identification of At-Risk Students: Using Learning Management System Data. *Journal of Postsecondary Student Success*, *2*(4), 108–126. https://doi.org/10.33009/fsop_jpss132082

Plak, S., Cornelisz, I., Meeter, M., & Van Klaveren, C. (2021). Early warning systems for more effective student counselling in higher education: Evidence from a Dutch field experiment. *Higher Education Quarterly*, *76*(1), 131–152. https://doi.org/10.1111/hequ.12298

Pursell, H. (2024, February 22). Mentor Matching: 7 Key Elements to Consider | Guider AI. Guider AI. Retrieved from https://guider-ai.com/blog/mentor-matching-key-elements/

Sholtis, S. (2022). Low grades in intro STEM courses may disproportionally deter minority students. Penn State University. Retrieved from https://www.psu.edu/news/eberly-college-science/story/low-grades-intro-stem-courses-may-disproportionally-deter-minority/#:~:text=%E2%80%94%20Receiving%20a%20grade%20of%20C,students%20with%20similar%20educational%20backgrounds

Sithole, A., Chiyaka, E. T., McCarthy, P., Mupinga, D. M., Bucklein, B. K., & Kibirige, J. (2017). Student Attraction, Persistence and Retention in STEM Programs: Successes and Continuing Challenges. *Higher Education Studies*, *7*(1), 46. https://doi.org/10.5539/hes.v7n1p46

Slavin, R. E., & Madden, N. A. (1989). What Works for Students at Risk: A Research Synthesis. *Educational Leadership*, *46*(5), 4–13. Retrieved from https://eric.ed.gov/?id=EJ383923

Su, Y. S., Lin, Y. D., & Liu, T. Q. (2022). Applying machine learning technologies to explore students' learning features and performance prediction. *Frontiers in Neuroscience*, 16. https://doi.org/10.3389/fnins.2022.1018005

Sundli, L. (2007). Mentoring—A new mantra for education?. *Teaching and Teacher Education*, *23*(2), 201–214. https://doi.org/10.1016/j.tate.2006.04.016

Xu, P., Ji, X., Li, M., & Lu, W. (2023). Small data machine learning in materials science. *NPJ Computational Materials*, *9*(1). https://doi.org/10.1038/s41524-023-01000-z

Yu, C. C., & Wu, Y. (2021). Early Warning System for Online STEM Learning—A Slimmer Approach Using Recurrent Neural Networks. *Sustainability*, *13*(22), 12461. https://doi.org/10.3390/su132212461

Zhao, W., Liu, X., Shah, S., Baah, I., Patel, A., & Wise, N. (2021). Peer support in smart learning and education. 2021 IEEE Conference Publication. *IEEE Xplore*. https://doi.org/10.1109/swc50871.2021.00089

Corresponding author: D. V. D. S. Abeysinghe
Email: dharani.21@cse.mrt.ac.lk