# Secure AI-Enhanced Student Engagement Analysis

TIMOTHY RUNHAAR, University of Twente, The Netherlands

This research explores the implementation of Artificial Intelligence (AI) in analysing and enhancing student engagement in university settings, focusing on the challenges of quantifying engagement and the potential of AI to provide objective insights. Traditional methods for assessing engagement, such as observations and surveys, lack consistency and real-time feedback. The emergence of AI, however, offers a solution by analysing data from student interactions within learning management systems. The study begins with a literature review on student engagement, identifying key influencing factors such as positive emotions towards learning, high-order learning, and student-staff interactions. The Open University Learning Analytics dataset (OULAD) is then analysed to extract patterns and indicators of engagement. The research utilises a Light Gradient Boosting Machine (LightGBM) model to predict student results based on their interactions with the online learning environment, revealing that higher interaction rates correlate with better academic results. A prototype web app is developed, showcasing a tool for educators to analyse student engagement data. The app features secure login mechanisms, interactive dashboards, and integrates machine learning predictions to identify at-risk students. Security and data privacy are emphasised, with measures such as JWT token access control and data encryption. The study acknowledges ethical considerations in using sensitive student data and the limitations of the OULAD dataset. Future work includes refining the machine learning model and gathering user feedback for tool improvement. In conclusion, the research demonstrates AI's potential in enhancing student engagement and academic success in online university settings, while emphasising the importance of ethical and security considerations in its application.

## 1 INTRODUCTION

The landscape of education is evolving rapidly, with a growing emphasis on technology integration and online learning[1]. Understanding and enhancing student engagement is crucial for effective teaching and improved learning outcomes [2]. However, quantifying engagement in campus-based higher education institutions is challenging due to the diverse forms it can take, such as lecture attendance, self-study, and usage of online/digital systems [3]. The difficulty lies in measuring this engagement accurately, especially in traditional face-to-face university environments, where student interactions with their learning programs and campus life are varied and numerous, necessitating innovative methods for their capture[4]. This research project aims to present the understandings of student engagement, analyse engagement data with correlation to achievements, leverage machine learning to build a predictive model using that data, and present a prototype of a web app to enhance student engagement in educational settings while maintaining the security and confidentiality of student data.


*TScIT 40, February 2, 2024, Enschede, The Netherlands*


## 2 PROBLEM STATEMENT

The need for students to be engaged in university settings comes without saying, there is found to be a moderately strong and positive correlation between overall student engagement and academic achievement[5]. In educational settings, particularly in universities, accurately measuring and enhancing student engagement remains a significant challenge. Traditional methods of assessing student engagement, such as teacher observations and student self-reported surveys, are often subjective and lack consistency[3]. This subjectivity can lead to a gap in understanding the true levels of student engagement, which is crucial for effective teaching and improved learning outcomes. Moreover, these traditional methods do not provide predictive analysis that can aid educators in promptly adapting their teaching strategies to meet students' needs. The emergence of AI offers a promising solution to these challenges. It is possible to analyse vast amounts of data from various student interactions within a learning management system, providing objective, data-driven insights into student engagement levels. However, the application of AI in this context does raise concerns regarding the security and confidentiality of sensitive student data. Therefore, this research project aims to develop a secure AI-enhanced tool designed for university settings. This tool provides a more objective and consistent analysis of student engagement but also incorporates security measures to ensure the protection of student data.

### 2.1 Research Questions

The problem statement leads to the following research question:

(1) How can AI be utilised to securely analyse student engagement in (online) university settings?
   (a) What factors influence student engagement in higher education?
   (b) What insights into student engagement can be revealed from the analysis of student interaction data with a Online Learning Environment ?
   (c) How can AI-driven analysis of student engagement be securely integrated into a web application prototype for educators?

## 3 METHODOLOGIES

### 3.1 Understanding student engagement

A comprehensive review of existing academic literature is conducted to identify and evaluate current approaches to understanding student engagement. This review focuses on the following key areas:

- Theoretical models for understanding student engagement
- Crucial factors for improving student engagement

### 3.2 Dataset Analysis and Machine Learning

The Open University Learning Analytics dataset (OULAD)[6] provided by the Knowledge Media Institute (KMI) at the Open University is analysed to extract meaningful patterns and identify potential engagement indicators. The OULAD dataset has been selected out of three datasets found publicly accessible. The two others being the KDDCup dataset[7] and the 365 Data Science Student Engagement Analysis dataset [8]. The former was not selected due to a lack of documentation. Whereas the latter dataset solely contained data within one year, insufficient for our use case. The chosen OULAD dataset consists of anonymised student interactions with the Virtual Learning Environment (VLE) combined with demographics and study results. Analysis is made on students' frequency of interactions, with what activity types they interact with and more, correlating it with student performances. The results give insights into what the most influential factors are.

For the machine learning(ML) part, the development uses a free and open-source distributed gradient-boosting framework for machine learning called Light Gradient Boosting Machine (LightGBM)[9] Gradient boosting is a machine learning technique proven to be able to accurately predict student academic performance achieving the highest accuracy among other techniques when used with the OULAD dataset[10]. We also consider this model for it's effectiveness when working with tabular data [11]. All students' interactions with learning materials are registered which correctly pre-processed should allow us to build an ML model to predict student's final results in a course. This involves the following steps:

- Data cleaning and pre-processing
- Exploratory Data Analysis (EDA), finding patterns
- Feature engineering
- Predictive modeling based on interaction data

### 3.3 Tool Development

The development of the secure AI-powered tool involves building a prototype web app. This showcases what an application providing insights based on student engagement data can look like. The tool provides features based on general findings on student engagement and data analysis from preceding sub-questions. Further the machine learning model offering predictive capabilities based on student engagement/interaction data is integrated. The tool is designed to meet the following requirements:

- Identifying engagement indicators and predicting student final result.
- Ease of use for educators
- Security measures, such as JSON Web Tokens(JWT) access control, data encryption, password hashing and salting.

## 4 WHAT FACTORS INFLUENCE STUDENT ENGAGEMENT IN HIGHER EDUCATION?

Student engagement in higher education is a multifaceted concept, it refers generally to the connection between a learner and their educational experiences. It involves the efforts of educational institutions to create programs and activities that cultivate a learning environment [12].

Engagement is challenging to quantify due to its diverse forms, like lecture attendance, self-study, and use of online/digital systems [3]. The paper by Ella R. Kahu [13] identifies the complexity of student engagement in higher education, emphasising the need for clearer definitions and distinctions between engagement, its influencing factors, and outcomes.

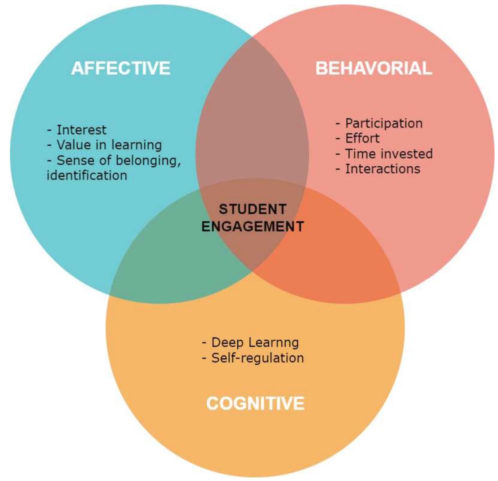



Fig. 1. Dimensions of student engagement

Figure 1 is a diagram built on the concept of the three dimensions of student engagement. Kahu finds that engagement is multidimensional and studies multiple proposed models and finds it is optimally split up in to three facets[13]. In this paper, we use the definition of these engagements as provided by Fredricks, J.A., McColskey, W. (2012) [3] that is in broad agreement with Kahu about the dimensions of student engagement. Describing behavioral engagement as the involvement in academic, social, or extracurricular activities. Affective/Emotional engagement as the interest, value, and reactions (positive or negative) to teachers, classmates, academics, or school. And lastly cognitive engagement as the self-regulation and learning strategies; student's investment in learning, including effort in comprehending complex ideas or mastering skills. The three facets can be presented like in Figure 1.

A study made by Lei, Hao & Cui, Yunhuo & Zhou, Wenye. (2018) [5] showcases the results of a meta-analysis, examining data from a number of independent studies on student engagement, in order to determine overall trends in the subject. It finds that there is a moderately strong and positive correlation between behavioral, emotional and cognitive engagement, and academic achievement. Furthermore, the study shows that behavioral engagement shows a slightly higher average effect size with academic achievement[5]. The paper includes a study on the correlation between each engagement facets and academic achievement. With the average weighted effect sizes (r) for the relationships found to be as follows:

(1) Overall Engagement: r = .269
(2) Behavioral Engagement: r = .350
(3) Emotional Engagement: r = .216
(4) Cognitive Engagement: r = .245

Additionally, interesting to take out of this study, is that for cognitive as well as emotional engagement, the correlations with academic achievement were higher among students from Eastern cultures compared to those from Western cultures. In contrast, for behavioral engagement the correlation with academic achievement was stronger in Western students than in Eastern students.[5].

However, in the aim of directing the development of the engagement analysis tool, it is important to identify key factors impacting student engagement. Through the literature review certain factors are found to be more critical than others. In no particular order, the first key factor found in students to translate into increased engagement, is a factor that is quite easy to assume. It is the student's positive emotions towards learning [13, 14]. The paper "Framing Student Engagement in Higher Education" by Ella R. Kahu [13] highlights that the affective dimension of engagement is an important aspect, which includes emotions such as enjoyment and interest in the learning task. This dimension contrasts instrumental motivation (engaging for external rewards like grades or qualifications) with intrinsic motivation, where the student is motivated by pleasure and interest in the learning itself. As another paper identifies individual positive emotion as one of the key promoting factors for student engagement[14].

Furthermore, a second factor translating significantly to engaged students, is high-order(deep) learning [13, 15, 20, 21]. It is defined as the involvement in advanced tasks like applying learned concepts, synthesising ideas, evaluating different sources of information, and creating new ideas[15]. The study by Ogunsakin (2021) [15] utilised Multiple Correspondence Analysis (MCA) to analyse student engagement data and establish weighty indicators of academic performance. The results showed that higher-order learning is one of two crucial indicator that relates student engagement to academic performance. Showing high-order learning to impact engagement and subsequently performances.

The last factor found to be crucial to high student engagement, is the interactions between students and staff [15]. The factor of positive student-teacher relationships is significantly more prominent than any other in studies about student engagement [3, 13–17]. As the paper by Jian Li and Eryong Xue(2023) says, positive and supportive interactions between students and teachers can significantly enhance student engagement and learning outcomes [14]. This paper suggests that positive teacher-student relationships foster a more engaging learning environment, thereby enhancing student performance. While on the other hand factors including lack of environmental support, negative student behavior, and negative teacher behavior have a hindering effect on student engagement. As another paper suggests, effective teacher-student relationships positively impact student engagement in academic activities and perceptions of workload [16]. Teachers and teaching practices have a significant impact on student engagement, relationships with staff are crucial to the learning situation and feeling part of a learning community positively influences student engagement[13].

Having identified key factors in student engagement in higher education, we must still know that students have different engagement typologies, making it difficult to generalise these factors impact on degree attainment in higher education [18]. Thus, finding a solution to increase each factor specifically remains a difficult task.

As previously mentioned, a notable count of papers emphasise the importance of the teacher-student relationship[3, 13–17]. Giving students a sense of value, a sense of purpose, is critical to positive emotion and engagement in students[20]. Using network analysis, the study by Korhonen (2019)[20] demonstrated that sense of belonging and identity were the most central components of engagement. Crucial as it also finds personal-intellectual motivation is positively associated with these exact two components[20]. Meaning that fostering a sense of belonging by creating a welcoming inclusive environment and supporting identity formation through for example career counseling and self-exploration opportunities will lead to motivation and positive emotion towards learning in students. It is also found that collaborative learning and peer interactions lead to positive student-student relationships [16], practices that can be seen as important in fostering positive emotions towards learning. Providing personalised learning experiences and timely feedback on assignments and assessments can help students understand their progress and areas for improvement[14]. This approach fosters a sense of individual attention and support[14], also leading to positive emotions in students.

Furthermore, providing personalised learning experiences and timely feedback on assignments and assessments also contribute to increase high-order learning in students. As the study by Korhonen (2019)[20] indicates a strong positive association between personal-intellectual motivation and the strategic approach to learning. This strategic approach is closely connected with study skills, which are essential for higher-order cognitive skills [20]. On the other hand, a surface approach to learning, characterised by a focus on memorisation and reproduction of information, is negatively associated with higher-order learning and is linked to trouble in finding appropriate study methods and intentions of dropping out [20]. Denoting that courses such as academic skills are consequent to high-order learning, and should thus have high importance to students and educators.

Figure 2, summarising the multiple ways that these engagement factors can be optimised is seen above. It is constructed based on the previous paragraphs with as format example a figure presented in a paper by Chiu(2021)[19].

Engagement with students is crucial, and so is the manner in which they receive feedback. Positive impacts from this feedback are essential for effective learning. In this context, a study titled 'A Story-Driven Gamified Education on USB-Based Attack' by Rikkers and Sarmah(2023)[2] explored the potential of story-driven gamification. This study evaluated learning outcomes and discovered that the inclusion of story elements was perceived as significantly enhancing the efficacy of gamification. Another insight contributing to the quest to optimise engagement factors in educational settings.

Conclusively, the multifaceted concept of student engagement in higher education is extensively explored, focusing on its three primary facets: behavioral, emotional, and cognitive engagement. The synthesis of various studies reveals a complex yet coherent picture of how these facets interact and contribute to academic achievement. Key findings indicate that while each facet has its unique influence, there is a notable interconnection among them, emphasising the importance of a holistic approach to fostering student engagement. Positive emotions towards learning, high-order learning, and student-staff interactions emerged as critical factors influencing student engagement. The research highlights that these factors do not operate in isolation; rather, they are part of a dynamic and interrelated system. For instance, positive emotions can be enhanced through meaningful student-staff interactions and engaging

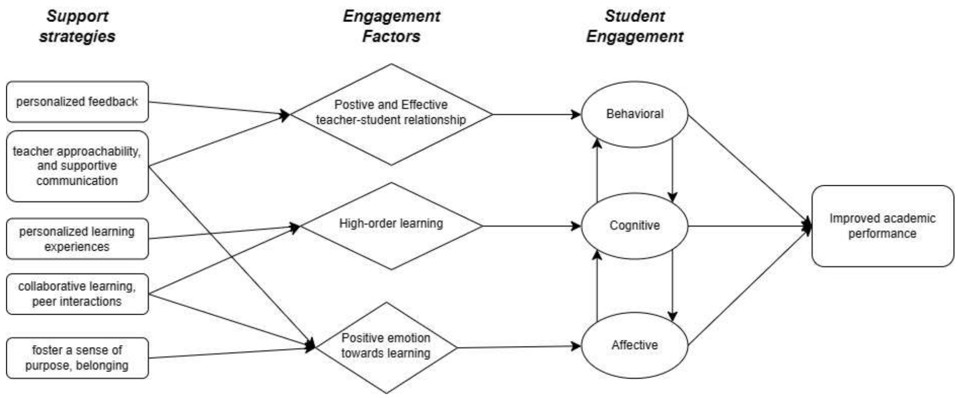



Fig. 2. Conceptual framework of engagement

in high-order learning activities. Similarly, effective student-staff relationships not only create a favorable learning environment but also motivate students to engage deeply with their academic pursuits. Furthermore, the significance of personal-intellectual motivation, sense of belonging, and identity in student engagement has been underscored, pointing towards the necessity of supportive and inclusive educational environments.

## 5 WHAT INSIGHTS INTO STUDENT ENGAGEMENT CAN BE REVEALED FROM THE ANALYSIS OF STUDENT INTERACTION DATA WITH A ONLINE LEARNING ENVIRONMENT ?

Having gained a better understanding of student engagement, we now know that behavioral engagement shows the most correlation with academic achievement. Meaning that participation, effort, time invested and interactions are crucial to student engagement and academic achievement. Knowing this we decide to analyse student's behavioral engagement with their online learning environment through a dataset logging student interactions. This allows us to further prove the importance of behavioral engagement, but also allow finding patterns relating interactions with achievement. As mentioned in the methodology, the OULAD dataset provided by the Open University was chosen to be the dataset to analyse. This dataset, notable for its good documentation, stands out as a rare public resource in the field of university learning analytics. thus making it an apt choice for this research. The objective now is to extract meaningful patterns and identify potential engagement indicators from this data. We start by explaining the structure of the dataset and proceed in performing an exploratory data analysis (EDA). In this section, we also explore the potential of machine learning, specifically through a Light Gradient Boosting Machine (LightGBM) model, in predicting student results based on their interactions with the online learning environment.

The first major step involves preparing the OULAD dataset for analysis. The dataset showed to contain already clean data, with the exception of some undefined values which were documented. Given the diverse nature of the data – encompassing student demographics, interaction logs, and assessment scores – restructuring the data proved to be a crucial step. To explain how this was done first should be understood how the OULAD dataset is structured. It is formed of data collected on 4 semesters, 2013B, 2013J, 2014B and 2014J(B for presentations starting in february and J for presentations starting October). There are several different modules (eg. FFF), which in combination with a presentation code(eg. 2014B) form a course(eg. FFF-2014B). Furthermore we notably want to analyse the data present on student's interactions with the learning material. Interaction data is logged in the dataset in the following structure inside the student_vle table: student id, module-presentation code, material id, the date relative to start of the course in days, and finally the amount of clicks made. Crucially the dataset also provides a final result per course for each student, indicating either Fail, Withdrawn, Pass or Distinction.

After initial analysis, the following insights were gained regarding the interactions (sum_click) of students with their Online Learning Environment. We categorised multiple features by final result and attempt to find the data showing the most significant distinctions, as this is the data that allows the ML model to perform classification with the most accuracy.

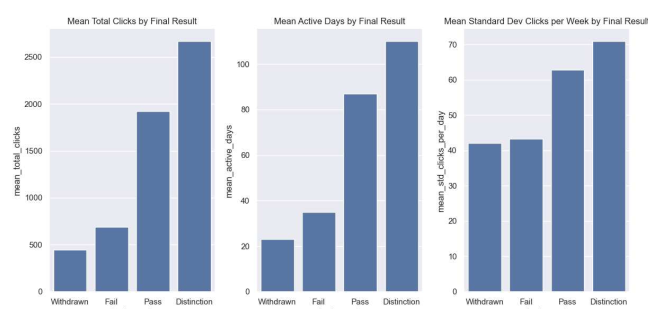



Fig. 3. Interaction data categorised by Final_result

It is observed that total clicks, active days, and the standard deviation of clicks per week measured in students provides a relatively clear distinction of classes. Essentially these two statistics proved to be the most different according to student's final results: Withdrawn: Students who withdrew from the course had an average of roughly 444 interactions, 23 active days. Fail: Students who failed had an average of around 688 interactions, 35 active days. Pass: Students who passed had an average of about 1922 interactions, 87 active days. Distinction: Students who achieved a distinction had an average of approximately 2667 interactions, 110 active days.

These charts suggest that students who are more engaged with the course materials (as indicated by clicks and active days) are more likely to pass. Higher interaction rates are associated with better final results (like passing and distinction), while lower interaction rates are observed in students who failed or withdrew.

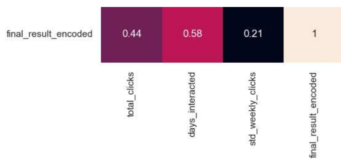



Fig. 4. Final_result correlation heatmap

To study this relationship more specifically, the correlation with the numerically encoded final results is computed using the Pearson[25] correlation method, shown in Figure 4. It is a statistical measure that evaluates the linear relationship between two variables. It ranges from -1 to +1, where +1 indicates a perfect positive linear relationship, 0 indicates no linear relationship and -1 indicates a perfect negative linear relationship. A correlation coefficient of approximately 0.438 is computed with the total number of VLE(Virtual Learning Environment) interactions(total clicks). While a correlation of 0.577 is computed with the amount of active days (days where at least one interaction has been made with the VLE). This suggests a moderate positive linear relationship, as the total clicks and active days increase, there is a tendency for the final result to be better (e.g., more likely to pass). For the standard deviation of weekly clicks a coefficient of 0.210 is calculated, suggesting a less strong relationship with final result.

It's important to note that correlation does not imply causation. This correlation indicates a relationship but does not definitively indicate that higher VLE interaction causes better results, or vice versa. But nevertheless it proves them potential to be useful features for our LightGBM classifier.

To further explore the potential predictive capabilities offered by interaction data, the aim is to build a machine learning model. To train the model the data needs to be structured with feature columns, which we will expand on next, and target column being the final result. Feature columns encompass the data the model uses to predict the target column. Building these feature columns is done by firstly merging data student interactions with student information. We group the data by student id, module code, and course code.

The interaction data with material in the VLE can be reconstructed to be classed by activity types(quizzes, urls, pdfs, forums, and more). As this feature had potential to feed more data to the model and improve classification, exploration is done. Interaction counts are grouped by activity type and categorised by final result. The chart in Figure 5 indicates that interactions grouped by activity types are not found to consistently relate to student success, not

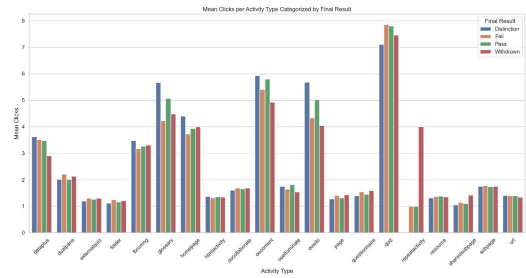



Fig. 5. Mean Clicks per Activity Type Categorised by Final Result

showing clear distinction per final result. Thus activity types are found to not be indicative of failing or withdrawing students in this dataset, and a pivot to a simpler dataframe had to be made to ensure an ML model could be delivered for this study.

So the previous findings in the analysis are taken and computations are performed to obtain the total number of clicks a student made in a module-presentation, the total number of days a student interacted with any course material, and the standard deviation of clicks per week(reflecting the variability in the student's weekly interactions). We exclude student id's as to not train the model on student's final result based on their id. This would allow for enhanced performances on the same dataset, but lead to overfitting and reduce performance of the model on unknown student ids. Producing the data structure shown in Figure 6.

| *Constructed dataframe* |
|---|
| code_module: Code name of the module. |
| code_presentation: Code for the presentation |
| total_clicks: The total number of clicks a student made |
| days_interacted: The total number of days a student interacted with the materia |
| std_weekly_clicks: The standard deviation of clicks per week, reflecting the variability in the student's daily interaction. |
| final_result: Final result in the course. (Withdrawn, Fail, Pass, or Distinction) |

Fig. 6. Data structure for ML

The model is trained using Optuna[22], a hyperparameter optimisation framework. This framework automatically tries to find the optimal parameters for the LightGBM model to improve its performance. By combining LightGBM with Optuna, we fine-tune to our specific dataset and prediction task. Once it is done running a 100 trials, we save the best parameters, the model, and its evaluations to analyse further. This allowed for the building of a model with the following evaluations: The obtained precision metric tells us how

```
Precision: 0.5845766492557563
Recall: 0.6057133082449538
F1 Score: 0.5450363238814909
```

Fig. 7. LightGBM overall evaluation

many of the positively predicted results (across all classes) were actually positive. A weighted average precision of 58.46% suggests that there's a fair amount of false positives in the predictions. The recall metric is a measure of the model's ability to find all the relevant cases within a dataset. The recall of 60.57% indicates that the model is moderately successful in identifying the positive cases. Meaning that on average for each class, the model correctly identified 60% of the the students part of that class. The F1 Score is a balance between precision and recall, providing a holistic view of the model's performance. An F1 score of 54.50% indicates that there is a balance between precision and recall, but it's not particularly high, suggesting the model could be improved; by means of more extensive feature engineering for example. An evaluation of the model per class is also allows us to understand the model's performance better.

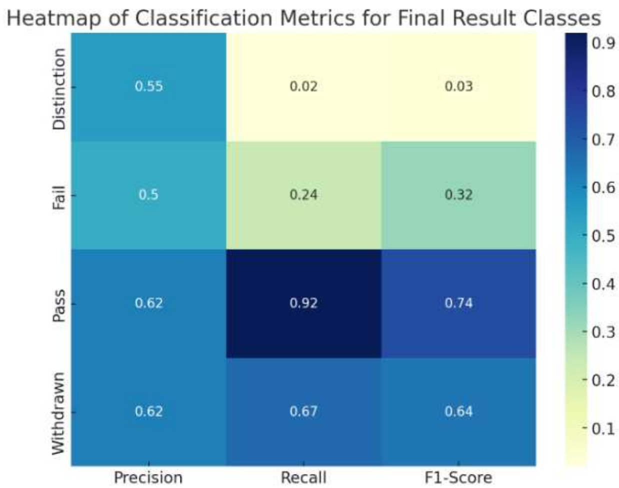



Fig. 8. LightGBM per class evaluation

The classification report provides a detailed view of the performance for each class ('Distinction', 'Fail', 'Pass', 'Withdrawn'). The model performs best in the 'Pass' category with a high precision (0.62) and recall (0.92), indicating reliable prediction of passing students. In contrast, the performance is weakest in the 'Distinction' category, with very low recall (0.02) despite moderate precision (0.55). The 'Withdrawn' category shows balanced precision and recall, resulting in a solid f1-score (0.64). And finally the model's performance in the 'Fail' category is moderate, with relatively lower precision and recall compared to 'Pass' and 'Withdrawn', reflected in its f1-score (0.32). In summary, what we find is the model performs well in identifying 'Pass' instances but struggles with the 'Distinction' category, showing low accuracy in identifying such cases, this is most likely due to tendency of the model to overfit to the majority class, which is here "Pass", and as such predicting students who will get "Distinction" to just get a "Pass". It exhibits moderate performance for 'Fail' and fairly good performance for 'Withdrawn' categories. What is interesting for the study is that the for all the students that have withdrawn the mode correctly identifies 67% of those students. This means that it misses only about 33% of the cases that should have been classified as "Withdrawn". A recall of 67% is fairly good, indicating that the model is relatively capable of

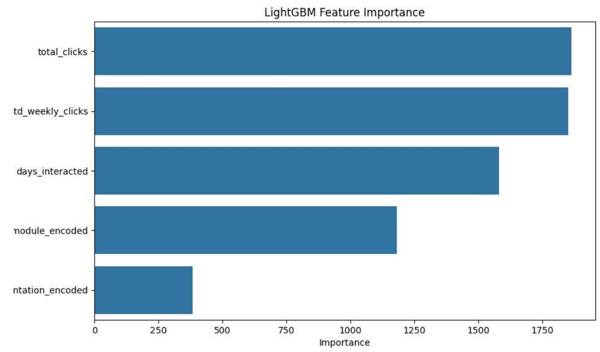



Fig. 9. Feature Importance

capturing the majority of "Withdrawn" cases. The varying performance across these categories suggests a need for model refinement and possibly addressing issues like imbalanced data or overfitting, especially to improve its ability to accurately classify 'Distinction' instances. We conclude with a feature importance analysis on the classifier

We perform a feature analysis to find what data the LightGBM model utilised the most in its task of making accurate classifications. This metric indicates the importance or contribution of each feature (input variable) in the prediction made by the model. The chart suggests that in our classifier, the total amount of clicks and the standard deviation of weekly clicks has more weight in predicting the final result, than the amount of days a student was active. Not conforming to the correlation analysis made previously where active days had the highest correlation coefficient. This is explained by the fact that active days and total clicks have a relatively high correlation coefficient of 0.81, and a feature highly correlated with the outcome may show lower importance in a model if other features capture similar information. The module_code showing a significant importance is surprising but definitely explainable. As it simply is an indication that some modules are easier or more difficult than others, thus allowing predictions on results to be made depending on which module a student is in.

In summary, the analysis finds that higher interaction rates, measured by total clicks, active days and standard deviation of clicks per week, are associated with better results. These features, found to be most distinct per final result, are leveraged in building of the LightGBM model. This data contributes significantly to predict student's final results up to a certain accuracy level. As suggests the analysis of feature importance, confirming the findings in the initial analysis of the dataset. This proves that with proper extraction and processing of interaction data, timely identification of at-risk (Fail or Withdraw) students deems possible. Within the limitations that come with the OULAD dataset; eg. limited data amount, unclear activity importance.

## 6 HOW CAN WE INTEGRATE THESE FINDINGS AND PRESENT A SECURE PROTOTYPE WEB APP OFFERING ANALYSIS OF STUDENT ENGAGEMENT DATA FOR EDUCATORS?

Based on the previous sections, this third part focuses on the development of a prototype web app that utilises the insights gained from the literature review and the machine learning analysis of the OULAD dataset. This web-app aims to provide educators with a tool to analyse student engagement in online university settings. This in an effort to conceptualise what a student engagement analysis tool could resemble. Here's a breakdown of the web app's features:

### 6.1 Login/Register Page (Figure 11)

Educators can log in with their credentials or register for a new account. Registration initiates an account with no course access initially. Secure authentication mechanisms are implemented to protect user credentials and ensure data privacy.

### 6.2 Dashboard (Figure 10)

This is the page educators are directed to after a successful login:

- **Course Selection**: A dropdown list enables educators to select a course from those they have access to.
- **Interactions per Day**: Displays the daily interactions of all students within a course. This feature allows educators to observe engagement trends and identify periods of reduced or increased student activity.
- **Predictive Results Distribution**: A pie chart presents the distribution of students' final results as predicted by the machine learning model. This visualisation aids educators in quickly assessing the overall performance and potential risks in the class.
- **Student List with Predictions and Details**: Students are listed with their predicted final results, prioritising at-risk students. Additional details per student include assessments(Figure 12), an interaction history chart(Figure 13) and submissions(Figure 14). This detailed view empowers educators to make informed decisions and offer targeted support to students in need.

### 6.3 Machine Learning Integration

When a course is selected, the list of student id's sent to the backend and used to generate predictions using a pre-built LightGBM model, based on each students interaction data with the Virtual Learning Environment (VLE). These predictions facilitate educators task of identifying students who may be at risk of failing or withdrawing.

### 6.4 Security and Data Privacy

- Secure Login/Register: Passwords are hashed and salted using bcrypt.
- JWT Token Access Control: Ensures that the dashboard and API calls are accessible only with a valid token.
- Student Information Access Control: Permissions are verified through JWT tokens, ensuring educators access only the data they are authorised to view. This is cross-checked against a course access table, ensuring stringent data access governance.
- Data Encryption: The Credentials and Course access datatables are encrypted using Fernet (symmetric encryption).

### 6.5 WebApp Technical Implementation

The web app's backend is developed in Python using the Flask framework, chosen for its flexibility and ease of use for creating web applications. The frontend is built using React.js, selected for its efficiency in rendering dynamic user interfaces. A local copy of the prototype tool can be ran by following the instructions in the git repositories' README.txt files. [23, 24]

## 7 ETHICAL AND SECURITY CONSIDERATIONS

On a ethical standpoint we choose to not utilise features in the OULAD dataset regarding socio-economic status, geographical data, disabilities or gender to build our machine learning model. They would have provided improved performances to the model but I deemed the use of these unethical. Additionally, despite proposed security measures, the dynamic nature of cybersecurity poses an ongoing challenge, with potential unforeseen vulnerabilities that could impact data confidentiality.

## 8 LIMITATIONS AND FUTURE WORK

A lot of time was dedicated to deepening the understanding of machine learning concepts and building the model, as there was not much experience in doing so. This caused a deviation from the initial planning, not leaving enough time for feedback collection on the tool. The findings are limited to the specific characteristics of the OULAD dataset. These are potentially restricting the generalisation to other educational contexts, as the usage of their VLE is specific to their courses and platform. This suggests the need to go through the process of preparing the data, building and evaluating models from scratch for use on Canvas data for example. Future enhancements include further refining the machine learning model for more accurate predictions. Additionally, user feedback will be integral to the app's development, ensuring it meets educators' needs effectively.

## 9 CONCLUSION

This research project successfully integrated artificial intelligence to analyse student engagement in online university settings. The literature review illuminated key factors influencing engagement, such as positive emotions, high-order learning, and student-staff interactions. The analysis of the OULAD dataset via a LightGBM model demonstrated a moderate ability to predict student outcomes based on interaction data, highlighting the potential of AI in identifying at-risk students. And finally the prototype web app developed exemplifies how these insights can be operationalised to assist educators. However, it's crucial to consider ethical and security aspects in handling sensitive student data. The project underscores the transformative potential of AI in enhancing student engagement and academic success, while also highlighting the need for continuous refinement and ethical consideration in its application.

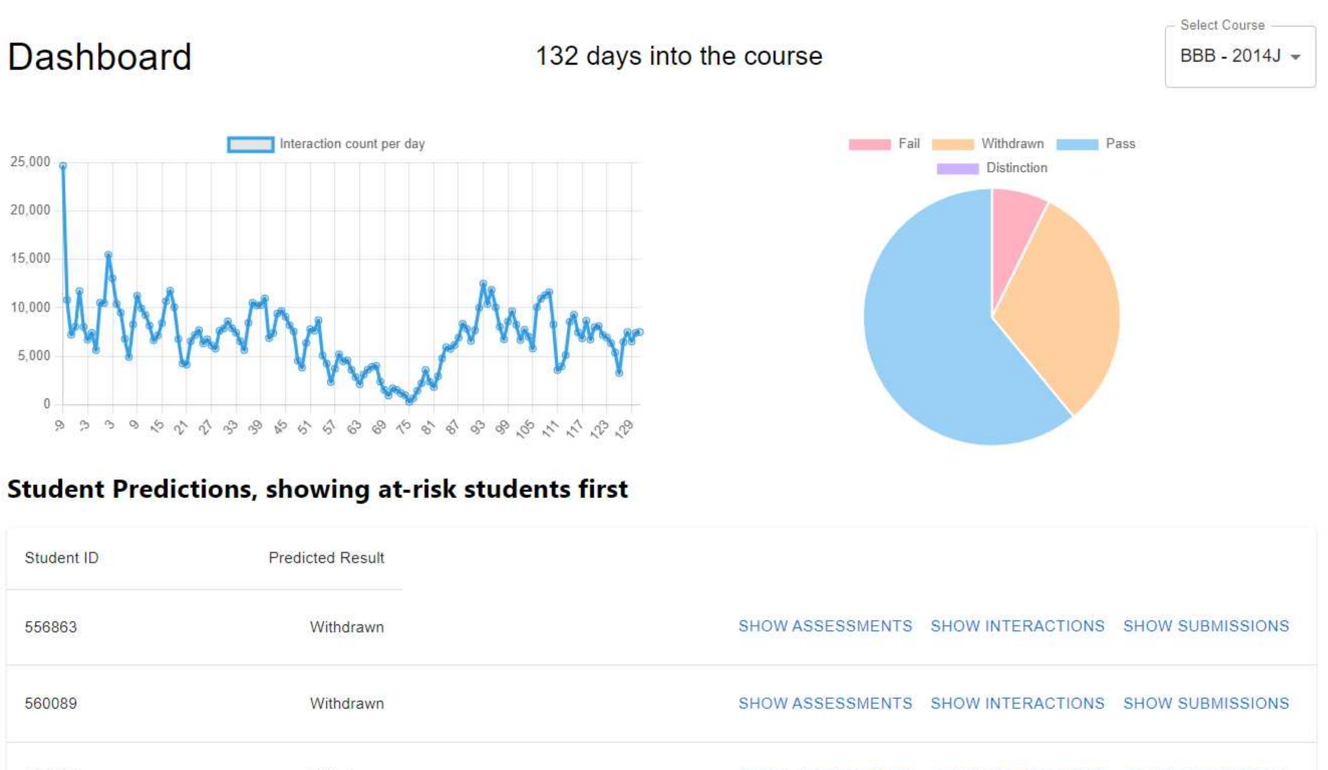



Fig. 10. Course Analytics Dashboard

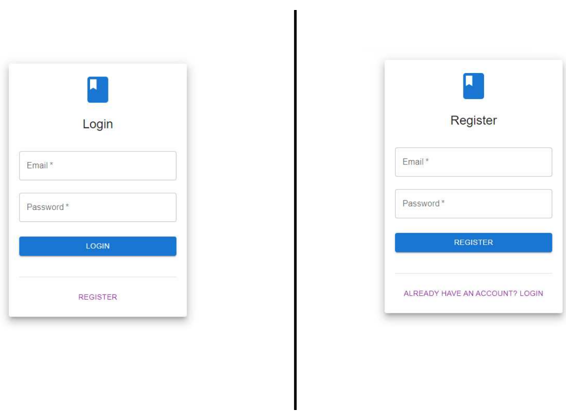
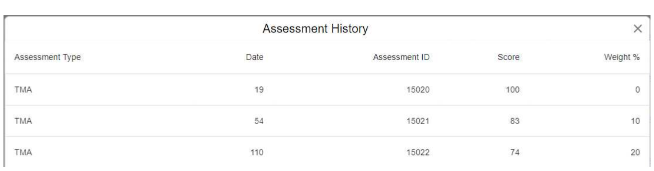



Fig. 11. Login page & Register page



Assessment History

| Assessment Type | Date | Assessment ID | Score | Weight % |
|---|---|---|---|---|
| TMA | 19 | 15020 | 100 | 0 |
| TMA | 54 | 15021 | 83 | 10 |
| TMA | 110 | 15022 | 74 | 20 |

Fig. 12. Student assessments pop-up

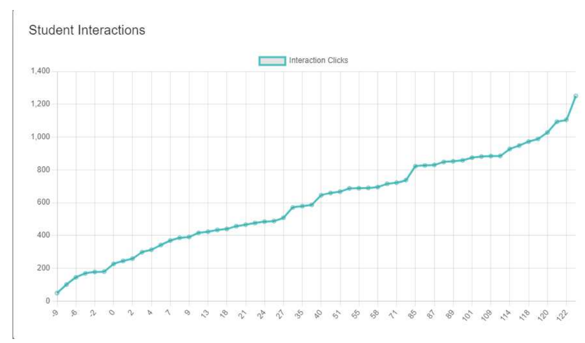
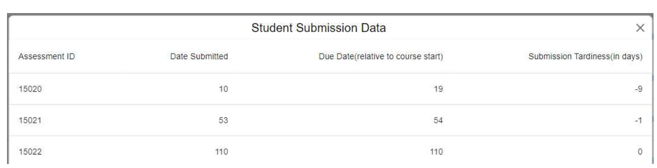



Fig. 13. Student interaction count evolution pop-up



Student Submission Data

| Assessment ID | Date Submitted | Due Date(relative to course start) | Submission Tardiness(in days) |
|---|---|---|---|
| 15020 | 10 | 19 | -9 |
| 15021 | 53 | 54 | -1 |
| 15022 | 110 | 110 | 0 |

Fig. 14. Student submissions pop-up

## REFERENCES


[1] Meng, N., Dhimolea, T.K., Ali, Z. (2022). AI-Enhanced Education: Teaching and Learning Reimagined. In: Albert, M.V., Lin, L., Spector, M.J., Dunn, L.S. (eds) Bridging Human Intelligence and Artificial Intelligence. Educational Communications and Technology: Issues and Innovations. Springer, Cham. https://doi.org/10.1007/978-3-030-84729-6

[2] Rikkers, V., Sarmah, D.K. A story-driven gamified education on USB-based attack. J Comput High Educ (2023). https://doi.org/10.1007/s12528-023-09392-z

[3] Fredricks, J.A., McColskey, W. (2012). The Measurement of Student Engagement: A Comparative Analysis of Various Methods and Student Self-report Instruments. In: Christenson, S., Reschly, A., Wylie, C. (eds) Handbook of Research on Student Engagement. Springer, Boston, MA. https://doi.org/10.1007/978-1-4614-2018-7_37

[4] Kim, J., Lee, H. & Cho, Y.H. Learning design to support student-AI collaboration: perspectives of leading teachers for AI in education. Educ Inf Technol 27, 6069–6104 (2022). https://doi.org/10.1007/s10639-021-10831-6

[5] Lei, Hao & Cui, Yunhuo & Zhou, Wenye. (2018). Relationships between student engagement and academic achievement: A meta-analysis. Social Behavior and Personality: an international journal. 46. 517-528. 10.2224/sbp.7054.

[6] https://analyse.kmi.open.ac.uk/open_dataset

[7] http://moocdata.cn/data/user-activity

[8] https://github.com/tripleaceme/365-Data-Science-Student-Engagement-Analysis

[9] https://lightgbm.readthedocs.io/en/stable/

[10] A. Merchant, N. Shenoy, A. Bharali and M. A. Kumar, "Predicting Students' Academic Performance in Virtual Learning Environment Using Machine Learning," 2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T), Raipur, India, 2022, pp. 1-6, doi: 10.1109/ICPC2T53885.2022.9777008.

[11] T. Sadasue and T. Isshiki, "Scalable Full Hardware Logic Architecture for Gradient Boosted Tree Training," 2020 IEEE 28th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM), Fayetteville, AR, USA, 2020, pp. 234-234, doi: 10.1109/FCCM48280.2020.00067.

[12] Goncalves, S., & Trunk, D. (2014). Obstacles to Success for the Nontraditional Student in Higher Education. Psi Chi Journal of Psychological Research, 19, 164-172. https://doi.org/10.24839/2164-8204.JN19.4.164.

[13] Ella R. Kahu e.r.kahu@massey.ac.nz (2013) Framing student engagement in higher education, Studies in Higher Education, 38:5, 758-773, DOI: 10.1080/03075079.2011.598505

[14] Li J, Xue E. Dynamic Interaction between Student Learning Behaviour and Learning Environment: Meta-Analysis of Student Engagement and Its Influencing Factors. Behavioral Sciences. 2023; 13(1):59. https://doi.org/10.3390/bs13010059

[15] Ogunsakin,R.,Moyo,S.,Oludayo,,Olugbara,O. & Israel,C.(2021).Relating Student Engagement Indicators to Academic Performance Using Multiple Correspondence Analysis. Cybernetics and Information Technologies,21(1) 87-102. https://doi.org/10.2478/cait-2021-0007

[16] Xerri, M.J., Radford, K. & Shacklock, K. Student engagement in academic activities: a social support perspective. High Educ 75, 589–605 (2018). https://doi.org/10.1007/s10734-017-0162-9

[17] Delfino, A.P. (2019). STUDENT ENGAGEMENT AND ACADEMIC PERFORMANCE OF STUDENTS OF PARTIDO STATE UNIVERSITY.

[18] Flynn, D. Baccalaureate Attainment of College Students at 4-Year Institutions as a Function of Student Engagement Behaviors: Social and Academic Student Engagement Behaviors Matter. Res High Educ 55, 467–493 (2014). https://doi-org.ezproxy2.utwente.nl/10.1007/s11162-013-9321-8

[19] Thomas K. F. Chiu (2022) Applying the self-determination theory (SDT) to explain student engagement in online learning during the COVID-19 pandemic, Journal of Research on Technology in Education, 54:sup1, S14-S30, DOI: 10.1080/15391523.2021.1891998

[20] Korhonen V, Mattsson M, Inkinen M and Toom A (2019) Understanding the Multidimensional Nature of Student Engagement During the First Year of Higher Education. Front. Psychol. 10:1056. doi: 10.3389/fpsyg.2019.01056

[21] Umbach, P.D., Wawrzynski, M.R. Faculty do Matter: The Role of College Faculty in Student Learning and Engagement. Res High Educ 46, 153–184 (2005). https://doi-org.ezproxy2.utwente.nl/10.1007/s11162-004-1598-1

[22] https://optuna.org/

[23] https://github.com/Timorun/EngBackend/

[24] https://github.com/Timorun/EngFrontend/

[25] (2008). Pearson's Correlation Coefficient. In: Kirch, W. (eds) Encyclopedia of Public Health. Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-5614-7_2569