

Predicting Academic Procrastination Among University Students Using Autoencoder Model

Anthony Kingsley Sackey*, Benjamin Appiah*, Adolph Sedem Yaw Adu *, Emmanuel Kwesi Baah**,
Eric Nana Agyekum Boamah***

**Department of Computer Science, Ho Technical University, Ho, Ghana*

***Department of Computer Science, Lancaster University Ghana*

****Department of Computer Science, Kwame Nkrumah University of Science & Technology, Kumasi, Ghana*

asackey@htu.edu.gh, bappiah@htu.edu.gh, aadu@htu.edu.gh, b.emmanuelkwesi@lancaster.edu.gh,
eanboamah@st.knust.edu.gh

Abstract— University students engaged in online learning often exhibit high levels of academic procrastination. Thus, various studies have investigated this phenomenon to provide solutions to mitigate its impact on academic work. However, these studies largely treated procrastination as a predictor of various outcomes, while other studies that treated procrastination as an outcome annotated the dataset manually. Investigating academic procrastination as an outcome gives a detailed insight to enable effective intervention for students, while manual annotation can be time-consuming and costly with large datasets. Other studies have also used machine learning algorithms to learn on clickstream, demographic, and assessment data to predict procrastination. The results of these studies show that they can be improved using other machine learning algorithms. This study, therefore, uses autoencoder, an artificial neural network, to uncover procrastination patterns (“no procrastinate”, “likelihood”, “procrastinate”) among students’ based on their clickstream, demographics, assessment, and admission data. In doing so, we leveraged the ability of autoencoders to extract meaningful data features and process data that can only be monitored in an online setting. Subsequently, the study uses a logistic regression algorithm to verify the result’s accuracy, precision, F1 score, and recall. The model performed optimally with a minimal precision of 99%. The findings also suggest that older students and students with high academic scores do not often procrastinate. In contrast, students who have taken a course severally often procrastinate.

Keyword— academic procrastination, autoencoder, clickstream, logistic regression

I. INTRODUCTION

ONCE limited to large desktop computers, the internet has permeated every aspect of human life. From smartphones in pockets to sensors in appliances, society is constantly engulfed in the invisible waves of the internet, seamlessly connecting and informing people. The internet is now the foundation of the interconnected world. With this internet advancement, educational institutions have leveraged its potency to convey education to students, thereby making online education an essential component of modern education. The worldwide outbreak of COVID-19 has also undeniably contributed to the growth and acceptance of online learning [1]. Online learning has resulted in the generation of massive amounts of digital data emanating from the digital footprint of both students and lecturers, which can be used to address some of the challenges confronting teaching and learning. On the contrary, the amount of social interaction between lecturers and students is significantly reduced in online learning. Students are required to demonstrate a high level of independence and self-control in online learning [2]. Further, lecturers supervise less of students’ learning activities in online learning, which, in turn, causes academic procrastination among students [3]. Academic procrastination is further deepened by students’ weak self-learning skills, insufficient self-management skills, and inadequate motivation to learn [3]. Online learning students are required to regulate their learning by managing their learning processes and regularly reviewing their study plans [4]. Further, students may be more exposed to social media and mobile phones in the context of online learning, which could increase their level of addiction to these platforms and increase their procrastination rate, negatively impacting their academic performance [5]. This is confirmed in yet another study, which indicates that there is a higher level of procrastination among students in online learning than in face-to-face learning [6]. Thus, it is significant that students’ procrastination is studied to provide solutions to mitigate the challenge.

To objectively study students’ procrastination, the use of

Manuscript received October 09, 2024. This manuscript is submitted to the invited journal as a follow-up to our paper entitled “Predicting Academic Procrastination Among University Students Using Autoencoder Model” accepted at the 27th International Conference on Advanced Communication Technology (ICACT2025).

Anthony Kingsley Sackey is a lecturer at the Department of Computer Science, Ho Technical University, Ho, Ghana. (corresponding author, phone: +233 249693372; e-mail: asackey@htu.edu.gh).

Benjamin Appiah is a lecturer at the Department of Computer Science, Ho Technical University, Ho, Ghana; e-mail: bappiah@htu.edu.gh.

Adolph Sedem Yaw Adu is a lecturer at the Department of Computer Science, Ho Technical University, Ho, Ghana; e-mail: aadu@htu.edu.gh.

Emmanuel Kwesi Baah is a lecturer at the Department of Computer Science, Lancaster University Ghana, Accra, Ghana; e-mail: b.emmanuelkwesi@lancaster.edu.gh.

Eric Nana Agyekum Boamah is a former student at the Department of Computer Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana; e-mail: eanboamah@st.knust.edu.gh.

machine learning algorithms to analyse students' clickstream data from a learning management system (LMS) instead of data from questionnaires has been recommended [7]. Clickstream data is a comprehensive and live recording of students' interactions from LMS [8] and can provide a more thorough understanding of students' behaviour [9]. Clickstream data is generated without interfering with the student's learning process. Further, large-scale, regular, and timely assessments of student behaviour can be generated from clickstream data [8]. On the other hand, clicks on important resources matter more than scattered clicks, as high-volume clicks might mean mindless surfing of the LMS interface. Clickstream data does not also consider external factors that might affect clicking, such as poor internet connectivity or students needing breaks. Previous studies [2], [10], [11] combined clickstream data with additional features such as demographic and assessment data to predict student academic procrastination. This stresses the importance of combining clickstream data with additional features, providing a richer dataset for procrastination prediction among students in an online setting. In these existing studies, where researchers combined clickstream data with demographics and assessment data to predict students' procrastination, it is noted that the performance of the algorithms used was relatively low, which requires further study to improve the performance [11]. Other studies used k-means algorithm to cluster the data [2], [10]. K-means demands the user to specify the number of clusters and may leave some points un-clustered [12], [13]. This study, therefore, uses clickstream data together with three other features: demographics, assessment, and admission data, in an attempt to predict students' procrastination behaviour. Autoencoder, an unsupervised learning algorithm, is used to learn these features to predict academic procrastination among students. Autoencoders are proposed in this study to avoid the adverse time and cost implications associated with manual data annotation. Our experimentation shows that the proposed model achieved an encouraging result in student procrastination.

The remainder of this work is structured as follows: related background, methodology, results and discussion, and conclusion.

II. RELATED BACKGROUND

A. Academic Procrastination

Academic procrastination involves irrational putting off or delaying students' completion of academic tasks [14]. It is common among university students [15],[16]. As also indicated, 48% of students always procrastinate, while 97% of students suffer the effect of procrastination [17]. Academic procrastination is known to have negative consequences, such as poor academic performance and school dropout [18]. It also negatively impacts students' physical and mental health [2]. Various studies have been conducted concerning academic procrastination to mitigate challenges associated with it. However, most existing studies treat procrastination as a predictor of an outcome such as academic performance [7]. This is confirmed by other studies [19]–[21]. Investigating academic procrastination as an outcome will

give a better insight into the complexities of procrastination to develop more targeted and effective interventions to help individual students overcome it.

B. Machine Learning for Procrastination Prediction

Machine learning techniques have recently gained ground in predicting students' academic procrastination. This is because there is great potential in using machine learning-based models to predict dilatory behaviour [7]. Machine learning algorithms can analyse huge datasets and identify complex patterns and relationships among variables. They can also continuously learn from new data and make improved predictions. Different approaches have been used to annotate datasets for machine learning predictions. For instance, human experts were used to annotate datasets, after which a logistic regression algorithm and artificial neural networks were used to make the predictions in a study [22]. Though the accuracy of the predictions was quite good (that is, Logistic regression produced an accuracy of 82% while artificial neural networks produced an accuracy of 83%), the use of human experts for the data annotation might have introduced some level of biases in the data. Further, this is not sustainable because, as the dataset increases, it will introduce huge cost and time implications, coupled with inconsistencies that different experts may introduce. This suggests that manually annotating the clickstream dataset for prediction will be extremely burdensome since the clickstream dataset from user interactions is usually huge.

To overcome the challenges of expert annotation, an unsupervised learning algorithm, K-means, has been used to cluster student behaviours into procrastinators and non-procrastinators [2]. This clustering enabled the authors to build a model for academic procrastination prediction, which achieved a good performance of 88.9% accuracy. K-means has also been used in clustering students' study behaviour into active learners, mild procrastinators, and severe procrastinators, after which students' procrastination tendency was predicted, giving good results [5]. It is known that the K-means algorithm is sensitive to outliers and initial centres of the clusters converges to the local minimum [23]. K-means also require the user to specify the number of clusters, making some data points un-clustered [12], [13]. These disadvantages reduce the performance of the K-means algorithm.

Contrastingly, autoencoder, a deep learning model, has demonstrated remarkable efficacy in tasks that do not require annotations. This unsupervised neural network algorithm is designed to reconstruct an output like the input, and its effectiveness has been highlighted [24]. For instance, an autoencoder was employed to predict mortality risk, achieving an impressive accuracy of 96.05% [25]. In another study, autoencoder-integrated models were compared to other deep learning models in predicting traffic flow, with the former showcasing superior performance [26]. These findings align with other studies, further underscoring the advantages of autoencoders [27]. This study, therefore, employs an autoencoder algorithm to discern procrastination patterns among university students based on their academic data. Subsequently, the study utilizes a logistic regression algorithm to forecast the likelihood of procrastination among

these students. To the best of the authors' knowledge, this study is the first to employ an autoencoder algorithm specifically to study academic procrastination among university students as also confirmed by recent studies [28][29].

III. METHODOLOGY

An autoencoder-based approach and a logistic regression classifier are proposed to predict academic procrastination tendencies in an online learning context. Initially, the dataset undergoes preprocessing, where numeric features related to student activities are transformed using scaling and one-hot encoding. An autoencoder neural network architecture is then defined and trained on the preprocessed data to learn a compressed representation of the input features, focusing on reducing dimensionality while preserving significant patterns. A logistic regression model is trained on the encoded features with pseudo-labels ('No Procrastinate', 'Likelihood', 'Procrastinate') derived from encoded features to predict procrastination behaviour.

A. Data

The dataset for the study is the Open University Learning Analytics Dataset, which encompasses detailed features of 32,593 students' profiles enrolled in 22 STEM courses in a

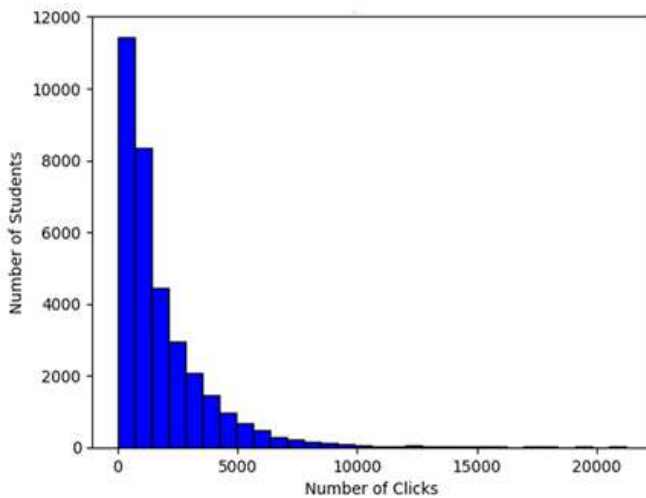


Fig. 1. Distribution of students' engagement (clickstream of students' interactions)

virtual learning environment. The dataset is distributed across multiple comma-separated values files to maintain accurate records for various education sector departments. It comprises multivariate sequential and time-series features, with student profiles spanning 2012 to 2014. This dataset contains students; assessment, admission, and a combination of demographic and aggregated clickstream data of students' interaction with the virtual learning environment. This facilitates the analysis of students' behaviour through their actions [28]. Fig. 1 shows the distribution of students' engagement (clickstream of students' interactions). The features of the dataset are shown in Table I.

TABLE I
FEATURES OF OPEN UNIVERSITY LEARNING ANALYTICS

S/No.	Features	Meaning of Features
1	gender	student's gender
2	region	the geographic region where the student lived while taking the module-presentation
3	highest_education	the highest student education level on entry to the module presentation
4	imd_band	the IMD band of the place where the student lived during the module-presentation
5	age_band	a band of student's age
6	num_of_prev_attempts	the number of how many times the student has attempted this module
7	studied_credits	the total number of credits for the modules the student is currently studying
8	disability	indicates whether the student has declared a disability
9	final_result	student's final result in the module-presentation
10	length	the length of the module presentation in days from the module start date to the module end date
11	date_registration	the day of the student's registration for the module presentation
12	date_unregistration	the day of student unregistration from the module presentation
13	assessment_type	a type of assessment. Three types of assessments exist tutor-marked assessment (TMA), computer-marked assessment (CMA), and final exam (Exam)
14	Ass_date	information about the cut-off day of the assessment
15	weight	the weight of the assessment
16	is_banked	the status flag indicating that the assessment result has been transferred from a previous presentation
17	score	the student's score in this assessment
18	date	the day of the student's interaction with the material
19	sum_click	the number of times the student interacted with the material
20	activity_type	the role associated with the module material
21	week_from	the week from which the material is planned to be used
22	Week_to	the week until which the material is planned to be used

B. Data Preprocessing and Extraction

As shown in Table I, the dataset for the study comprises 22 features. Among the features, gender, highest_education, region, imd_band, age_band and disability are categorical. One hot encoding was used to transform the categorical variables into binary vectors while the numerical features were scaled to a common scale using StandardScaler, which ensures that all the features have an equal influence on the model. The preprocessed data is then used to train the autoencoder model, which learns to compress the input data into a lower-dimensional representation while preserving relevant features.

C. Autoencoder for Academic Procrastination Prediction

Autoencoders (AE) have emerged as powerful tools in unsupervised learning, particularly for discovering latent patterns and extracting meaningful representations from complex data. In predicting academic procrastination among university students, autoencoders offer a unique approach to identifying procrastination-related patterns based on students' online interaction behaviour. By leveraging the architecture of autoencoders, it becomes possible to transform the intricate and multi-dimensional traits associated with procrastination tendencies into a lower-dimensional latent space. This transformation allows for the compression of the input features, effectively distilling the essential information that encapsulates procrastination behaviours.

A significant advantage of using autoencoders in unsupervised procrastination prediction is their ability to learn and encode patterns in the input data autonomously. In the case of student online interaction behaviour, this could encompass a wide range of variables, such as frequency of task delays, time spent on non-academic activities, and engagement with online educational resources. The autoencoder takes these diverse behavioural cues and compiles them into a compressed representation, which abstractly captures the underlying procrastination patterns. These patterns may include subtle yet significant associations between online behaviours and procrastination tendencies, shedding light on the complex interplay of factors contributing to academic procrastination. Autoencoder is composed of an encoder and a decoder. Each contains one or more layers of neurons/units. The encoder maps the input layer x to a lower-dimensional latent space representation y . The decoder mirrors the encoder, mapping the same layer y , reconstructing the data from the latent space to its output layer z . The learned code/representation consists of the layer y . When dealing with real values, training an AE involves optimizing the reconstruction error $\|x-z\|^2$.

D. Designed Autoencoder Model

The architecture of the proposed Autoencoder (AE) (as shown in Listing 1) is controlled by the number of layers, neurons per layer, and activation functions in the encoder and decoder. Fig. 2 shows the architecture of the autoencoder model used in this study. The encoder is represented by the variable input layer beginning with the first dense layer. This layer performs a linear transformation on the input data. It introduces non-linearity through the Rectified Linear Unit (ReLU) activation function, denoted as `activation='relu'`. The output of this first dense layer, referred to as `encoded`, serves as the intermediate representation. The encoder then proceeds to reduce the dimensionality through additional dense layers further, each with its set of weights and biases. The last dense layer in the encoder reduces the representation to a dimension specified by `encoding_dimencoding`, which is often significantly lower than the input dimension. This dimensionality reduction is achieved by applying the ReLU activation function in each layer.

On the other hand, the decoder is initialized with the same set of dense layers, including the last dense layer with an output dimension of `encoding_dimencoding`. This symmetric

design allows the decoder to mirror the encoding process. However, it is essential to note that the activation functions and layer dimensions match those of the encoder in the decoder. The decoder's output layer employs the sigmoid activation function suitable for the data's range.

```
#Encoder
encoded = Dense(256, activation='relu')(input)
encoded = Dense(128, activation='relu')(encoded)
encoded = Dense(64, activation='relu')(encoded)
#Decoder
decoded = Dense(128, activation='relu')(encoded)
decoded = Dense(256, activation='relu')(decoded)
decoded = Dense(2, activation='sigmoid')(decoded)
```

Listing 1: The AE shows the number of layers and dimensions of each layer. The "input" here is the academic online interaction records from students.

E. Training and Prediction

The training process of the autoencoder involves minimizing the Mean Squared Error (MSE) loss function, as expressed in (1). Where x represents the input data, and y is the reconstructed output from the autoencoder.

$$L(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2 \quad (1)$$

$$f(z) = \frac{1}{1 + e^{-z}} \quad (2)$$

During training, the autoencoder learns to encode the input data into a lower-dimensional latent space while striving to minimize the reconstruction error, effectively capturing meaningful features from the data. Once the autoencoder is trained, the learned latent representation (latent features) is input to a predictive model. Logistic regression, a widely used classification algorithm, is employed for prediction. Logistic regression utilizes the sigmoid function, denoted as (2) Where z represents the weighted sum of the learned latent features to model the probability of a binary outcome.

By leveraging the mathematical foundations of MSE loss and logistic regression, this approach seamlessly combines unsupervised feature learning with predictive modelling, enabling the identification of procrastination patterns in student academic behaviour.

IV. RESULTS AND DISCUSSION

A. Model Performance

We assessed the performance of the autoencoder model implemented based on the f1-score, precision, accuracy, and recall metrics as evidenced in other studies on academic performance from literature [30]. However, some studies employed a qualitative approach to this evaluation [31]. The model was evaluated on the Open University Learning Analytics dataset to validate and illustrate the model's efficiency in predicting academic procrastination. The model consistently performed adequately and optimally with a minimal precision score of 0.99 and 100% on the other specified metrics. Fig. 3 illustrates the model's performance. All the metrics used to assess the model are at their peak, except for precision, which negligibly falls short. The model in this study far outperforms other models in existing studies [5], [22], [29]. This demonstrates the robustness and

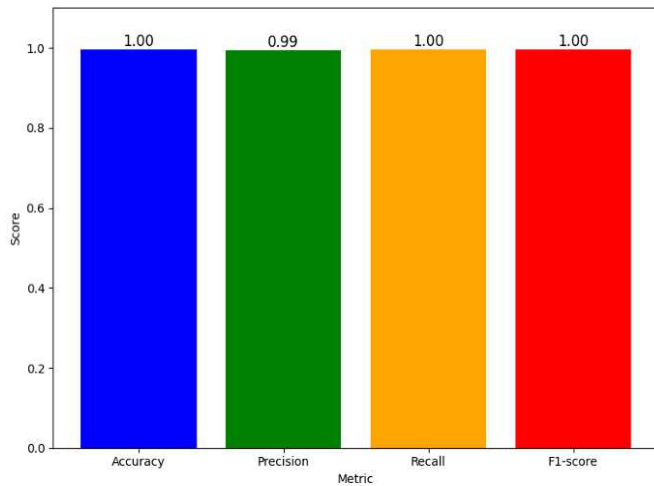


Fig. 3. Evaluation metrics to validate and illustrate the autoencoder model's efficiency in predicting academic procrastination

dependability of the model in the present study to correctly classify the data instances.

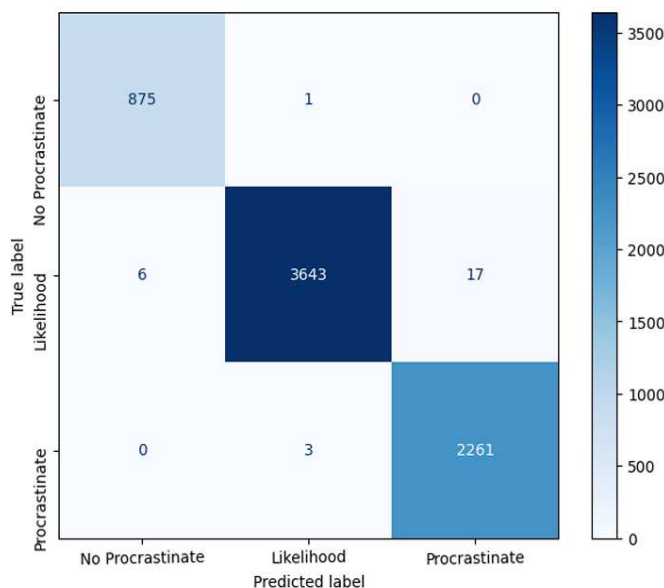


Fig. 4. Logistic Regression Confusion Matrix values

A further illustration of the performance of the autoencoder model on the test data is expressed in the confusion matrix, as shown in Fig. 4. The result indicates a multi-class problem involving 3 classes, namely “No Procrastinate”, “Likelihood” and “Procrastinate”. The model correctly predicted almost all (875) data instances as “No Procrastinate” and misclassified one (1) instance as belonging to the “Likelihood” class. It correctly predicted most of the data instances (3643) as the “Likelihood” class. However, it misclassified 6 and 17 of them under the “No Procrastinate” and “Procrastinate”

classes, respectively. As befits its reliability, it rightly placed almost all (2261) of the encountered test data instances under the “Procrastinate” class and only missed a few (3).

The silhouette score measures the distance between each data point in each cluster and the other data points in its neighbouring clusters. It is a useful tool for assessing and validating the quality of clusters produced by the model. Three clusters are visualized with colours, each signifying a prediction class. The mapping of the yellow, deep purple, and green clusters corresponds to the “No Procrastinate”, “Likelihood” and “Procrastinate” classes, respectively. Two factors were considered when assessing the quality of clusters and computing the silhouette score: cohesion and separation. Separation measures the distance between data points from different clusters.

In contrast, cohesion measures the distance between data points inside the same cluster or class. The value for cohesion must be small (highly cohesive) but large for separation. With very few outliers, most data points in each cluster in Fig. 5 are accurately categorized or associated with their neighbours (cohesion). It is evident from the graph that the different kinds of data points and clusters are becoming more distant from one another (separation), indicating the degree to which the model could distinguish between these classes. Generally, Fig. 5 demonstrates that these clusters are of good quality,

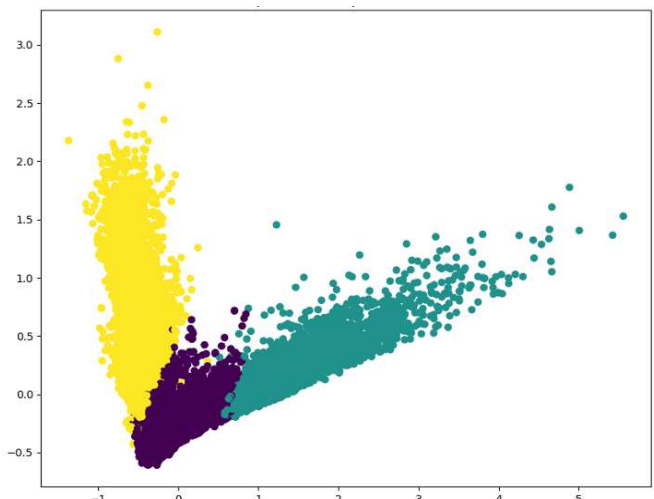


Fig. 5. 2D Representation space for the three clusters based on autoencoder

throwing more light on the prediction efficiency of the proposed autoencoder model.

B. Feature Importance in Academic Procrastination

In academic procrastination prediction, understanding the contribution of each feature to the overall prediction is key. Thus, the study examined how the Open University Learning Analytics Dataset features influenced or contributed to the autoencoder model's decisions, as shown in Fig. 6. The comparatively strong and weak contributors, as well as their impact on the weight value that produced the model's prediction, are shown in Fig. 6. The graph indicates the relative importance of each feature or variable in the model's decisions and shows a relationship among them. Fig. 6 shows

that ten (10) variables significantly impact academic procrastination prediction. It can be inferred that the score had the most significant impact on academic procrastination, suggesting that a lower score indicates academic procrastination-related issues. Student's highest education follows this. It also suggests that students pursuing higher education might be more familiar with managing complex tasks and meeting deadlines, which can potentially lead to lower procrastination. This aligns with the study of Özer [32], where it was found that undergraduate students procrastinate more than graduate students. It can also be inferred that students who have retaken a course several times tend to procrastinate more. This suggests that students who have failed a course may feel less motivated or lack confidence in their abilities, leading to procrastination. Previous negative experiences or underlying skill deficiencies may cause the students to be apprehensive, causing them to procrastinate.

Fig. 7 shows a correlation analysis between clickstream (representing student online activities) and the ten (10) important features that contribute to students' academic procrastination. This provides deeper insight into how these

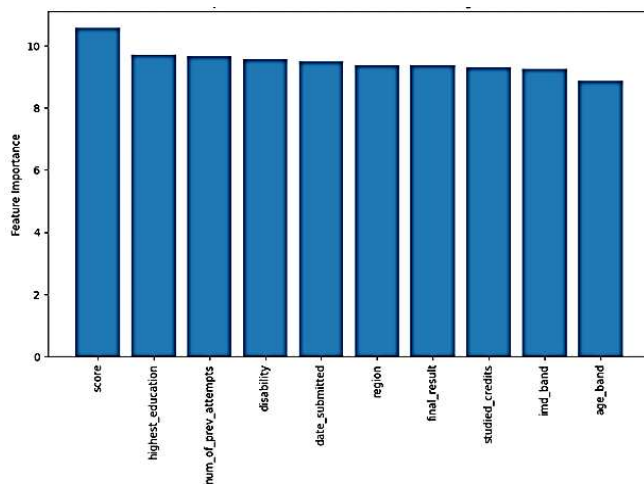


Fig. 6. Feature importance in procrastination prediction among students

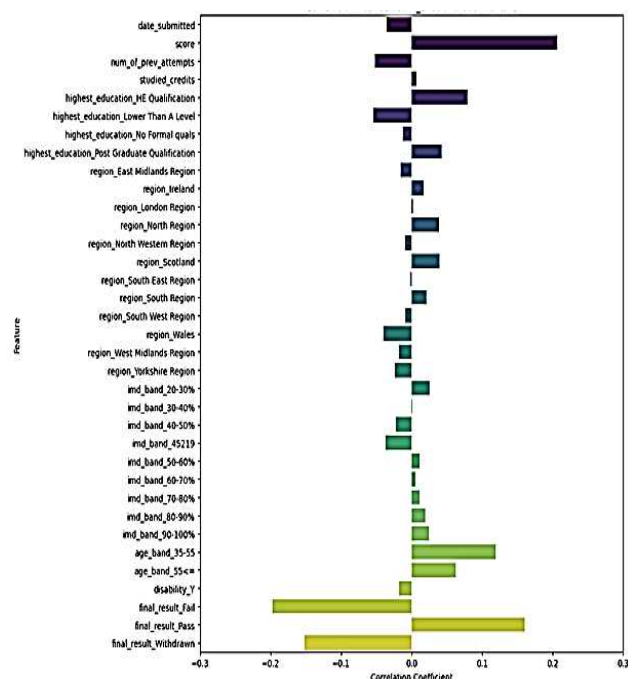


Fig. 7. Correlation matrix between students' clickstream and each feature

factors relate to procrastination in an online learning environment.

Engagement and Performance

The correlation results show that “score” has a moderately positive correlation (0.207) with “sum_click” (clickstream). This suggests that students who engage more online tend to achieve higher scores. This positive correlation aligns with the idea that active participation and interaction with course materials can enhance understanding and performance [33].

Conversely, “final_result_Fail” exhibits a strong negative correlation (-0.198) with “sum_click”, indicating that students who fail are less likely to have engaged actively online. This negative relationship underscores the importance of consistent online engagement in academic success. This finding is consistent with an existing study that indicates that students who actively engage in academic activities rarely procrastinate [34].

Socio-economic and Demographic Factors

Among the demographic factors, “age_band_35-55” shows a notable positive correlation (0.119) with “sum_click”, indicating that older students tend to engage more online than their younger counterparts. This finding suggests that older students, possibly with more responsibilities and a clearer educational goal, prioritize online activities more consistently. Regarding educational background, “highest_education_HE Qualification” has a positive correlation (0.080), implying that students with higher education qualifications tend to engage more online. Conversely, “highest_education_Lower Than A Level” exhibits a negative correlation (-0.055), suggesting less engagement among students with lower educational qualifications.

Socio-economic and Demographic Factors

Geographically, some regions show varying degrees of correlation with “sum_click”. For instance, “region_Scotland” (0.040) and “region_North Region” (0.039) exhibit positive correlations, indicating higher engagement levels in these areas. Conversely, “region_Wales” (-0.040) and “region_West Midlands Region” (-0.018) show negative correlations, suggesting lower engagement in these regions. These differences could reflect regional variations in access to resources, educational culture, or other socioeconomic factors influencing online learning engagement.

The correlation analysis highlights that online engagement (“sum_click”) in educational activities is positively associated with academic performance (“score”) and certain demographic factors like age and educational background. Conversely, there are negative associations with factors such as failure outcomes (“final_result_Fail”) and some geographic regions (“region_Wales”, “region_West Midlands Region”). These findings underscore the complex interplay of individual, socioeconomic, and institutional factors in shaping student engagement and outcomes in online learning environments, providing valuable insights for

interventions to reduce procrastination and improve educational outcomes.

V. CONCLUSION

The study sought to use students' online engagement data (clickstream data) together with demographics, assessment, and admission data to cluster students into "No Procrastinate", "Likelihood Procrastinate", and "Procrastinate". An unsupervised learning algorithm, autoencoder, was used to learn the dataset's features. A logistic regression algorithm was afterward used to verify the model's performance. This was done to overcome the time implications of manual annotation and challenges related to the K-means clustering algorithm. The model was assessed on the f1-score, precision, accuracy, and recall metrics to validate and illustrate the model's efficiency in predicting academic procrastination. The model consistently performed adequately and optimally with a minimal precision score of 0.99 and 100% on the other specified metrics. It also emerged from the study that less procrastination is associated with older students compared to younger ones; students who have retaken a course severally procrastinate more, while high levels of procrastination among students in general lead to poor academic performance. Training and experimenting with more current and extensive datasets are sure ways of improving the reliability of our work in the future. Additionally, exploring more lightweight prediction models like the Random Forest, Decision Tree, and other algorithms on such current datasets could yield valuable insights into predicting academic procrastination among the new generation of university students.

REFERENCES

- [1] A. Oginni, E. Mogaji, and P. N. Nguyen, *Re-imagining Educational Futures in Developing Countries: Lessons from Global Health Crises*, no. June. 2022.
- [2] H. Xu, J. Qu, and Y. Ling, "Prediction and Visualization of Academic Procrastination in Online Learning," pp. 133–139, 2021, <https://doi.org/10.1145/3474995.3475017>.
- [3] Y.-T. Ling, J.-H. Qu, X. Ma, and H.-M. Xu, "Research on Learner Academic Procrastination Based on Multi-view Clustering," pp. 106–113, 2021, <https://doi.org/10.1145/3474995.3475014>.
- [4] I. Roll and P. H. Winne, "Understanding, evaluating, and supporting self-regulated learning using learning analytics," *J. Learn. Anal.*, vol. 2, no. 1, pp. 7–12, 2015, <https://doi.org/10.18608/jla.2015.21.2>.
- [5] F. Feng, M. Tang, and W. Lei, "Study of Procrastination in Higher Vocational Education Based on Online Learning Data," *ACM Int. Conf. Proceeding Ser.*, pp. 82–87, 2022, <https://doi.org/10.1145/3568739.3568755>.
- [6] M. De Paola, F. Gioia, and V. Scoppa, "Online teaching, procrastination and student achievement," *Econ. Educ. Rev.*, vol. 94, no. February, p. 102378, 2023, <https://doi.org/10.1016/j.econedurev.2023.102378>.
- [7] C. Imhof, I. S. Comsa, M. Hlosta, B. Parsaeifard, I. Moser, and P. Bergamin, "Prediction of Dilatory Behavior in eLearning: A Comparison of Multiple Machine Learning Models," *IEEE Trans. Learn. Technol.*, vol. 16, no. 5, pp. 648–663, 2023, <https://doi.org/10.1109/TLT.2022.3221495>.
- [8] Q. Li, R. Baker, and M. Warschauer, "Using clickstream data to measure, understand, and support self-regulated learning in online courses," *Internet High. Educ.*, vol. 45, no. January, p. 100727, 2020, <https://doi.org/10.1016/j.iheduc.2020.100727>.
- [9] R. Baker *et al.*, "The benefits and caveats of using clickstream data to understand student self-regulatory behaviors: opening the black box of learning processes," *Int. J. Educ. Technol. High. Educ.*, vol. 17, no. 1, pp. 1–24, 2020, <https://doi.org/10.1186/s41239-020-00187-1>.
- [10] F. Rodriguez, H. R. Lee, T. Rutherford, C. Fischer, E. Potma, and M. Warschauer, "Using clickstream data mining techniques to understand and support first-generation college students in an online chemistry course," *ACM Int. Conf. Proceeding Ser.*, pp. 313–322, 2021, <https://doi.org/10.1145/3448139.3448169>.
- [11] N. S. Raj and V. G. Renumol, "An Approach for Early Prediction of Academic Procrastination in e-Learning Environment," *Int. J. Inf. Educ. Technol.*, vol. 13, no. 1, pp. 73–81, 2023, <https://doi.org/10.18178/ijiet.2023.13.1.1782>.
- [12] B. Chen, P. C. Tai, R. Harrison, and Y. Pan, "Novel hybrid hierarchical-K-means clustering method (H-K-means) for microarray analysis," *2005 IEEE Comput. Syst. Bioinforma. Conf. Work. Poster Abstr.*, pp. 105–108, 2005, <https://doi.org/10.1109/CSBW.2005.98>.
- [13] A. Bansal, M. Sharma, and S. Goel, "Improved K-mean Clustering Algorithm for Prediction Analysis using Classification Technique in Data Mining," *Int. J. Comput. Appl.*, vol. 157, no. 6, pp. 35–40, 2017, <https://doi.org/10.5120/ijca2017912719>.
- [14] B. U. Özer, A. Demir, and J. R. Ferrari, "Exploring academic procrastination among turkish students: Possible gender differences in prevalence and reasons," *J. Soc. Psychol.*, vol. 149, no. 2, pp. 241–257, 2009, <https://doi.org/10.3200/SOCP.149.2.241-257>.
- [15] Y. Fentaw, B. T. Moges, and S. M. Ismail, "Academic Procrastination Behavior among Public University Students," *Educ. Res. Int.*, vol. 2022, 2022, <https://doi.org/10.1155/2022/1277866>.
- [16] P. Steel, "The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure," *Psychol. Bull.*, vol. 133, no. 1, pp. 65–94, 2007, <https://doi.org/10.1037/0033-2909.133.1.65>.
- [17] S. He, "A Multivariate Investigation into Academic Procrastination of University Students," *Open J. Soc. Sci.*, vol. 05, no. 10, pp. 12–24, 2017, <https://doi.org/10.4236/jss.2017.510002>.
- [18] W. Doherty, "An analysis of multiple factors affecting retention in Web-based community college courses," *Internet High. Educ.*, vol. 9, no. 4, pp. 245–255, 2006, <https://doi.org/10.1016/j.iheduc.2006.08.004>.
- [19] A. Gareau, M. Chamandy, K. Kljajic, and P. Gaudreau, "The detrimental effect of academic procrastination on subsequent grades: the mediating role of coping over and above past achievement and working memory capacity," *Anxiety, Stress Coping*, vol. 32, no. 2, pp. 141–154, 2019, <https://doi.org/10.1080/10615806.2018.1543763>.
- [20] M. Goroshit, "Academic procrastination and academic performance: An initial basis for intervention," *J. Prev. Interv. Community*, vol. 46, no. 2, pp. 131–142, 2018, <https://doi.org/10.1080/10852352.2016.1198157>.

- [21] M. Goroshit and M. Hen, "Academic procrastination and academic performance: Do learning disabilities matter?," *Curr. Psychol.*, vol. 40, no. 5, pp. 2490–2498, 2021, <https://doi.org/10.1007/s12144-019-00183-3>.
- [22] M. M. Memon, M. A. Hashmani, S. M. Jameel, S. Junejo, and K. Raza, "Analysis of Student Procrastinatory Behavior in Virtual Learning Environments Using Machine Learning," *J. Hunan Univ. (Natural Sci.)*, vol. 47, no. 10, pp. 139–152, 2020, [Online]. Available: <http://jonuns.com/index.php/journal/article/view/459/456>.
- [23] S. E. Y. Nayini, S. Geravand, and A. Maroosi, "A novel threshold-based clustering method to solve K-means weaknesses," *2017 Int. Conf. Energy, Commun. Data Anal. Soft Comput. ICECDS 2017*, pp. 47–52, 2018, <https://doi.org/10.1109/ICECDS.2017.8389496>.
- [24] S. J. Kim, W. Y. Jo, and T. Shon, "APAD: Autoencoder-based Payload Anomaly Detection for industrial IoT," *Appl. Soft Comput. J.*, vol. 88, p. 106017, 2020, <https://doi.org/10.1016/j.asoc.2019.106017>.
- [25] F. Khozimeh *et al.*, "Combining a convolutional neural network with autoencoders to predict the survival chance of COVID-19 patients," *Sci. Rep.*, vol. 11, no. 1, pp. 1–18, 2021, <https://doi.org/10.1038/s41598-021-93543-8>.
- [26] W. Wei, H. Wu, and H. Ma, "An autoencoder and LSTM-based traffic flow prediction method," *Sensors (Switzerland)*, vol. 19, no. 13, pp. 1–16, 2019, <https://doi.org/10.3390/s19132946>.
- [27] M. Iatrou, C. Karydas, X. Tseni, and S. Mourelatos, "Representation Learning with a Variational Autoencoder for Predicting Nitrogen Requirement in Rice," *Remote Sens.*, vol. 14, no. 23, 2022, <https://doi.org/10.3390/rs14235978>.
- [28] D. Xu, F. Zhu, Q. Liu, and P. Zhao, "Improving exploration efficiency of deep reinforcement learning through samples produced by generative model," *Expert Syst. Appl.*, vol. 185, Dec. 2021, <https://doi.org/10.1016/j.eswa.2021.115680>.
- [29] Y. Lin, H. Chen, W. Xia, F. Lin, Z. Wang, and Y. Liu, "A Comprehensive Survey on Deep Learning Techniques in Educational Data Mining," pp. 1–21, 2023, [Online]. Available: <http://arxiv.org/abs/2309.04761>.
- [30] Y. Yang, D. Hooshyar, M. Pedaste, M. Wang, Y. M. Huang, and H. Lim, "Prediction of students' procrastination behaviour through their submission behavioural pattern in online learning," *J. Ambient Intell. Humaniz. Comput.*, no. 0123456789, p. 12652, 2020, <https://doi.org/10.1007/s12652-020-02041-8>.
- [31] R. Fang and R. White, "Distance Learning Students' Perspectives of Academic Procrastination: A Qualitative Investigation," *J. Contin. High. Educ.*, vol. 0, no. 0, pp. 1–17, 2024, <https://doi.org/10.1080/07377363.2024.2317614>.
- [32] B. U. Özer, "A Cross Sectional Study on Procrastination: Who Procrastinate More?," *Int. Conf. Educ. Res. Innov.*, vol. 18, pp. 34–37, 2011.
- [33] C. Vizoso, C. Rodríguez, and O. Arias-Gundín, "Coping, academic engagement and performance in university students," *High. Educ. Res. Dev.*, vol. 37, no. 7, pp. 1515–1529, 2018, <https://doi.org/10.1080/07294360.2018.1504006>.
- [34] Y. L. San, S. B. Roslan, and F. Sabouripour, "Relationship between self-regulated learning and academic procrastination," *Am. J. Appl. Sci.*, vol. 13, no. 4, pp. 459–466, 2016, <https://doi.org/10.3844/ajassp.2016.459.466>.



interests include Machine Learning, Educational Technology, Human Computer Interaction and Learning Analytics.



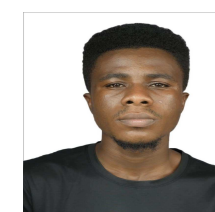
Benjamin Appiah is a Ghanaian born in 1986. He received his master's and PhD degrees in Software Engineering from University of Science & Technology of China. He is currently a lecturer at the Department of Computer Science, Ho Technical University, Ghana where he has been working since 2021. His research interests include machine learning, Security, Deep Learning, Data Mining, and Big Data Analysis.



Usability and User Experience evaluation and e-Learning.



security, and business intelligence.



Eric Agyekum Nana Boamah is a Ghanaian who was born in 1999. He graduated from Kwame Nkrumah University of Science and Technology (KNUST), Ghana in 2022 where he pursued B.Sc. Computer Science. He has been working as a private software developer and currently with SmartTransyt as a Field Technician. His areas of interest are computer software development and practical applications of machine learning.