Does Time Matter in Analyzing Educational Data? - A New Dataset for Streaming Learning Analytics

Gabriella Casalino[1,*], Giovanna Castellano[1] and Gianluca Zaza[1,*]

[1]*Computer Science Department, University of Bari Aldo Moro Bari, Italy*

Abstract

This research introduces a novel dataset developed for streaming learning analytics, derived from the Open University Learning Analytics Dataset (OULAD). The dataset incorporates essential temporal information that captures the timing of student interactions with the Virtual Learning Environment (VLE). By integrating these time-based interactions, the dataset enhances the capabilities of stream algorithms, which are particularly well-suited for real-time monitoring and analysis of student learning behaviors. Experiments utilizing the Online Bagging algorithm across three temporal units—months, trimesters, and semesters—demonstrated that the dataset contains pertinent information for predicting student outcomes. Despite the variations associated with different temporal units, the classifier effectively identified patterns within the data, especially for the majority class (Pass), achieving high F1 scores. These results indicate that the temporal structure of the data supports accurate predictions; however, challenges remain in accurately identifying the minority class (Fail). This dataset paves the way for more dynamic and responsive educational interventions by enabling timely predictions of student outcomes. Such capabilities facilitate continuous learning support within VLEs, allowing educators to respond promptly to student needs and enhance overall learning experiences.

Keywords

Learning Analytics, Dataset, Education, Machine Learning, Stream Data Mining

1. Introduction

Learning Analytics (LA) represents a specialized domain within artificial intelligence dedicated to the analysis of educational data, particularly focusing on students' interactions with Virtual Learning Environments (VLE). These interactions consistently generate significant data, including login timestamps, engagement metrics with course materials, participation in discussions, and assignment submissions [1]. A thorough analysis of this data can yield valuable insights for all stakeholders involved in the educational process [2], enabling them to understand student behaviors, identify learning patterns, and monitor academic progress in real time [3]. The continuous generation of such data is vital for implementing timely interventions aimed at enhancing the learning experience. Students can leverage this information to pinpoint areas for improvement and refine their study habits, while educators can utilize these insights to enhance

AIxEDU - 2nd International Workshop on Artificial Intelligence Systems in Education - 25-28 November 2024, Bolzano, Italy

*Corresponding author.

gabriella.casalino@uniba.it (G. Casalino); giovanna.castellano@uniba.it (G. Castellano); gianluca.zaza@uniba.it (G. Zaza)

000-0003-0713-2260 (G. Casalino); 0000-0002-6489-8628 (G. Castellano); 0000-0003-3272-9739 (G. Zaza)

course design. Furthermore, academic institutions can establish tailored support measures based on the analysis.

Nevertheless, the majority of existing literature tends to treat educational data as a static entity, failing to account for the inherently dynamic nature of the learning process [4, 5, 6, 7, 8]. An exception to this trend is found in studies focused on knowledge tracing, a machine-learning methodology that models and predicts the evolution of an individual student's learning trajectory over time through the analysis of their interactions with educational content. This approach utilizes temporal information to monitor the progression of a student's knowledge, thereby allowing for more accurate predictions of future performance based on historical learning behaviors [9, 10].

However, only a limited number of studies have integrated temporal factors into their analyses, apart from those concerning knowledge tracing. For instance, it has been observed in [11] that traditional machine-learning methodologies often overlook the dynamic nature of educational interactions, which could be more effectively analyzed through data stream methodologies. Stream algorithms are specifically designed to process continuous data in real time, efficiently managing large or continuous data streams with constrained memory and computational resources [12]. This capability renders them particularly suitable for analyzing student interactions with VLEs, given the substantial and ongoing data generated by student activities. The same authors also introduced explainable stream algorithms to facilitate the processing and interpretation of educational data, taking into account the temporal evolution of student interactions for more insightful analyses [13, 14]. Nevertheless, a significant obstacle impeding the progress of learning analytics for evolving data remains the scarcity of time-stamped data.

In this study, we utilized the information contained in the well-established Open University Learning Analytics Dataset (OULAD) [15] to create a student-oriented dataset that preserves the temporal evolution of user interactions, in conjunction with other demographic and evaluative information. Our dataset comprises 1 718 984 entries, with each instance characterized by 34 features that delineate the number of student interactions with each VLE facility (e.g., forum, glossary, homepage, HTML activity, etc.), alongside demographic data and the student's grade, for a specific student, course, and time T. The target classes are classified as follows: Withdraw, Fail, Pass, and Distinction.

We employed a stream classification algorithm, specifically Online Bagging, applied to data partitioned into smaller subsets based on temporal units such as months, trimesters, and semesters. This approach was designed to ascertain whether the resultant data provides sufficient information to predict students' outcomes over time.

The primary research question we seek to address is: " (Does time influence the effectiveness of classification models in predicting students' outcomes?" To explore this question, we examined various temporal periods, commencing with six months, followed by three months, and finally one month. We utilized the Online Bagging stream classification algorithm to investigate: i) whether the information within these temporal frames is adequate for accurate predictions; ii) whether there are performance differences based on the chosen periods; and iii) whether the stream classifier can effectively learn from small, incremental data subsets.

The structure of the manuscript is organized as follows: Section 2 delineates the methodology employed to extract temporal information from the OULAD dataset, along with a comprehensive description of the newly constructed data and relevant statistics. Section 3 elaborates on the

experimental procedures involving the stream algorithm, demonstrating how to process the new data and assess classification performance. Finally, Section 4 summarizes our conclusions and outlines potential avenues for future research.

2. Data

One significant limitation associated with the application of stream algorithms within the educational context is the scarcity of data containing temporal information. To address this challenge, we manipulated the Open University Learning Analytics Dataset (OULAD) [15] to incorporate time-based data, thereby enabling us to leverage the timing of student interactions. However, it is important to note that the anonymization process employed in the original OULAD dataset means that the temporal information is not absolute; rather, it is relative to the start date of each module, which remains undisclosed.

The primary files utilized from the OULAD dataset in this study include:

- courses.csv: Contains information regarding course modules and their presentations, including columns such as ‘code_module‘, ‘code_presentation‘, and ‘length‘.
- assessment.csv : Details the assessments corresponding to each module presentation, with columns including ‘code_module‘, ‘code_presentation‘, ‘id_assessment‘, ‘assessment_type‘, ‘date‘, and ‘weight‘.
- vle.csv : Provides information on the materials available within the VLE, featuring columns such as ‘id_site‘, ‘code_module‘, ‘code_presentation‘, and ‘activity_type‘.
- studentInfo.csv : Contains demographic data and student results, encompassing columns such as ‘code_module‘, ‘code_presentation‘, ‘id_student‘, ‘gender‘, ‘region‘, ‘highest_education‘, ‘imd_band‘, ‘age_band‘, ‘num_of_prev_attempts‘, ‘studied_credits‘, ‘disability‘, and ‘final_result‘.
- studentRegistration.csv : Details student registrations for modules, including ‘code_module‘, ‘code_presentation‘, ‘id_student‘, ‘date_registration‘, and ‘date_unregistration‘.
- studentAssessment.csv : Records student assessment results, with columns such as ‘id_assessment‘, ‘id_student‘, ‘date_submitted‘, ‘is_banked‘, and ‘score‘.
- studentVle.csv : Captures student interactions with VLE materials, featuring ‘code_module‘, ‘code_presentation‘, ‘id_student‘, ‘id_site‘, ‘date‘, and ‘sum_click‘.

To construct a comprehensive dataset, we initially merged data from various CSV files to consolidate information pertaining to modules and presentations.

In detail, the initial step involved merging the student_info and student_vle tables based on the features code_module , code_presentation , and id_student , resulting in a consolidated table that integrated information from both sources. Subsequently, this consolidated table was merged with the vle table, utilizing code_module , code_presentation , and id_site as reference points to create a table containing information on activity types. The assessments and student_assessment tables were merged using the id_assessment feature. Finally, the two resulting tables were joined using the id_student and date features, thereby creating a comprehensive dataset that encapsulates relevant student interactions and assessment results.

We then analyzed student interactions with VLE materials to identify usage patterns and assess engagement levels. Additionally, we integrated assessment results with interaction dates to examine the correlation between student interactions and performance outcomes. The dataset underwent a thorough cleaning and transformation process to address missing values and normalize the data, ensuring consistency and reliability in our analysis.

Upon consolidating the data into a unified table, which includes student information and their interactions for a specific student S, course C, module M, and a given time T (measured in days from the start date of each module), we sought to transform this temporal information. The original dataset recorded a minimum temporal granularity of one day, thus precluding the extraction of day partitions. Given that the dataset provides only relative time values, our objective was to convert these into a linear and incremental format spanning the years represented in the OULAD data. This transformation allows us to maintain the relative nature of the time while ensuring usability for analysis and effectively reflecting the temporal progression of student interactions throughout their studies.

Modules in the original OULAD data can commence in February (denoted by the letter B) or October (denoted by the letter J). To ensure consistency and establish a progressive timeline from a single reference date, we transformed these dates into a count of incremental days, commencing from the first module in February (B). The transformation process involved the following steps:

1. Identification of the Base Date : We established the start date for the 2013B presentation as the reference point (zero).
2. Calculation of the Effective Date : We converted the relative dates of the other presentations into days elapsed since the identified start date, taking into account the number of days between the various modules.

This methodology enabled us to create a unified temporal framework conducive to analysis, accurately reflecting the progression of module timelines.

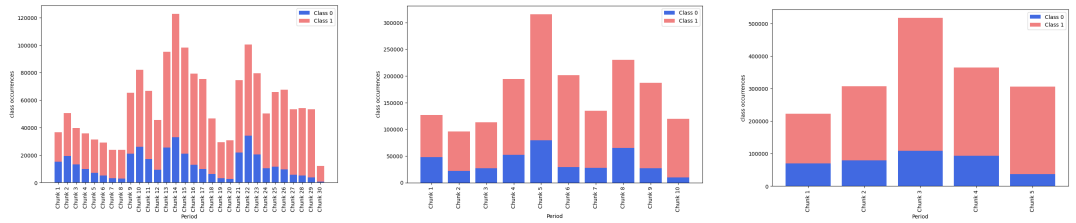
As a result, we obtained a dataset[1] consisting of 34 features and 1 718 983 samples, encompassing students' demographic information, assessment scores, and interactions with the VLE for a specific time T, corresponding to each student S within a given course C and module M. Table 1 provides a comprehensive description of all features[2]. Additionally, categorical features were converted to numerical representations, with mappings included for reference in the last column. The target classes—'Withdrawn', 'Fail', 'Pass', and 'Distinction'—were encoded as 0, 1, 2, and 3, respectively. It is noteworthy that the data exhibits a significant imbalance, with a substantial prevalence of records associated with students who passed the final examination. The class distribution is as follows: 'Pass' (1 022 760 samples), 'Distinction' (308 642 samples), 'Fail' (227 550 samples), and 'Withdrawn' (160 031 samples) [16].

[1] Casalino, G., Castellano, G., Zaza, G. (2024). A New Dataset for Streaming Learning Analytics [Data set]. Zenodo. https://doi.org/10.5281/zenodo.14003233

[2] Please note that the features 'dualpanel', 'folder', 'repeat activity', and 'html activity' were excluded from the dataset due to the absence of recorded interactions.

Table 1
Description of the features of the new dataset with the corresponding mapping.

Features	Description	Mapping
code_module	module identification code on which the student is registered	'AAA'=0; 'BBB'=1; 'CCC'=2; 'DDD'=3; 'EEE'=4; 'FFF'=5; 'GGG'=6
code_presentation	presentation identification code during which the student is registered on the module	'2013B'=0; '2013J'=1; '2014B'=2; '2014J'=3
id_student	the unique student identification number	-
gender	student's gender	M=0; F=1
region	the geographic region where the student lived while taking the module presentation	'East Anglian Region'=0; 'Scotland'=1; 'North Western Region'=2; 'South East Region'=3; 'West Midlands Region'=4; 'Wales'=5; 'South Region'=6, 'South West Region'=7, 'East Midlands Region'=8, 'Yorkshire Region'=9; 'London Region'=10; 'North Region'=11; 'Ireland'=12
highest_education	the highest student education level on entry to the module presentation	-
imd_band	the IMD band of the place where the student lived during the module presentation	'0-10%'=0; '10-20%'=1; '20-30%'=2; '30-40%'=3; '40-50%'=4; '50-60%'=5; '60-70%'=6; '70-80%'=8; '80-90%'=9; '90-100%'=10
age_band	a band of student's age	'0-35'=0; '35-55'=1; '55<='=2
num_of_prev_attempts	the number of how many times the student has attempted this module	-
studied_credits	the total number of credits the student has already earned	
disability	indicates whether the student has declared a disability	'N=0'; 'Y=1'
final_result	student's final result in the module presentation	-
date	the date of student's interaction with the VLE measured as the number of days since the start of the module presentation	-
dataplus	the date of the student's interaction with the VLE in the linear time framework we created	
dualpane	Number of occurrences for the activity *dualpane* extracted from the *activity_type* feature of the *vle* table from the original *oulad* dataset	-
externalquiz	Number of occurrences for the activity *externalquiz* extracted from the *activity_type* feature of the *vle* table from the original *oulad* dataset	-
folder	Number of occurrences for the activity *folder* extracted from the *activity_type* feature of the *vle* table from the original *oulad* dataset	-
forumng	Number of occurrences for the activity *forumng* extracted from the *activity_type* feature of the *vle* table from the original *oulad* dataset	-
glossary	Number of occurrences for the activity *glossary* extracted from the *activity_type* feature of the *vle* table from the original *oulad* dataset	-
homepage	Number of occurrences for the activity *homepage* extracted from the *activity_type* feature of the *vle* table from the original *oulad* dataset	-
htmlactivity	Number of occurrences for the activity *htmlactivity* extracted from the *activity_type* feature of the *vle* table from the original *oulad* dataset	-
oucollaborate	Number of occurrences for the activity *oucollaborate* extracted from the *activity_type* feature of the *vle* table from the original *oulad* dataset	-
oucontent	Number of occurrences for the activity *oucontent* extracted from the *activity_type* feature of the *vle* table from the original *oulad* dataset	-
ouelluminate	Number of occurrences for the activity *ouelluminate* extracted from the *activity_type* feature of the *vle* table from the original *oulad* dataset	-
ouwiki	Number of occurrences for the activity *ouwiki* extracted from the *activity_type* feature of the *vle* table from the original *oulad* dataset	-
page	Number of occurrences for the activity *page* extracted from the *activity_type* feature of the *vle* table from the original *oulad* dataset	-
questionnaire	Number of occurrences for the activity *questionnaire* extracted from the *activity_type* feature of the *vle* table from the original *oulad* dataset	-
quiz	Number of occurrences for the activity *quiz* extracted from the *activity_type* feature of the *vle* table from the original *oulad* dataset	-
repeatactivity	Number of occurrences for the activity *repeatactivity* extracted from the *activity_type* feature of the *vle* table from the original *oulad* dataset	-
resource	Number of occurrences for the activity *resource* extracted from the *activity_type* feature of the *vle* table from the original *oulad* dataset	-
sharedsubpage	Number of occurrences for the activity *sharedsubpage* extracted from the *activity_type* feature of the *vle* table from the original *oulad* dataset	-
subpage	Number of occurrences for the activity *subpage* extracted from the *activity_type* feature of the *vle* table from the original *oulad* dataset	-
url	Number of occurrences for the activity *url* extracted from the *activity_type* feature of the *vle* table from the original *oulad* dataset	-
score	the student's score in the intermediate assessments. The range is from 0 to 100.	



(A) (B) (C)

Figure 1: Classes distribution over the chunks for different time units: months (A), trimesters (B), semesters (C).

3. Experiments and Discussion

Experiments were conducted to evaluate the suitability of the new dataset for stream algorithms. To achieve this objective, three sets of experiments were designed, each utilizing different temporal units: months (M), trimesters (T), and semesters (S). The data was segmented into intervals based on these temporal units, with groupings established every 30 days (for months), 90 days (for trimesters), and 180 days (for semesters). This segmentation resulted in 30 groups for the monthly analysis, 10 for the trimester analysis, and 5 for the semester analysis.

Given that the target classes 'Withdrawn' and 'Distinction' are underrepresented in comparison to the more prevalent classes 'Fail' and 'Pass', we opted to consolidate these categories into a binary classification scheme[3]. Specifically, 'Fail' was combined with 'Withdrawn', while 'Pass' was merged with 'Distinction'. This approach resulted in a dataset comprising 1 331 402 total samples for the Pass class and 387 581 samples for the Fail class.

Figure 1 presents statistics regarding class distribution across the various time segments for the three experimental settings: months (A), trimesters (B), and semesters (C). In the figure, the Pass class is represented in red, while the minority class, Fail, is depicted in blue.

It is important to note that the size of the chunks is not uniform across the three settings; rather, the number of samples within each chunk is contingent upon the quantity of interactions recorded within a given time period. Furthermore, data variability tends to increase as the temporal unit T is reduced. For instance, in the semester setting, the ratio between the two classes remains relatively consistent across the five chunks. Conversely, in the trimester setting, we observe variations in the distributions of the two classes across the different chunks. This variability becomes even more pronounced in the monthly data, where Fail cases are significantly outnumbered by Pass cases in some chunks.

Temporal patterns emerge from the analysis, revealing periods characterized by elevated levels of interaction, which correspond to higher incidences of both Failures and Successes. Following these peaks, the numbers tend to decline over subsequent months before experiencing another resurgence.

The Online Bagging algorithm was employed in this study due to its efficacy in managing

[3]The following mapping was adopted for the binary classification: Mapping = {'Withdrawn': 0, 'Fail': 0, 'Pass': 1, 'Distinction': 1}. In this scheme, 'Withdrawn' and 'Fail' were both mapped to 0 (indicating the Fail class), while 'Pass' and 'Distinction' were mapped to 1 (indicating the Pass class).

Table 2
Average classification performances and standard deviations over the chunks in different periods: Months (M), Trimesters (T), and Semesters(S), obtained with the Online Bagging stream classifier.

TU	Accuracy	Precision		Recall		F1	
		Fail	Pass	Fail	Pass	Fail	Pass
M	67.93 ± 1.75	37.8 ± 7.9	73.51 ± 3.37	20.92 ± 1.92	86.41 ± 0.67	26.86 ± 3.55	79.39 ± 1.78
T	69.22 ± 0.54	38.62 ± 4.35	75.09 ± 1.5	22.88 ± 1.21	86.46 ± 0.45	28.72 ± 2.13	80.36 ± 0.62
S	73.8 ± 1.36	42.66 ± 6.13	76.6 ± 1.78	14.31 ± 2.30	93.63 ± 0.16	21.43 ± 3.34	84.25 ± 1.13

evolving data streams through the application of ensemble methods, which enhance predictive performance and robustness in dynamic environments compared to other stream algorithms [17]. Online Bagging is an ensemble learning technique that adapts the conventional Bagging method for streaming data. Unlike traditional Bagging, which utilizes multiple bootstrap samples derived from a static dataset, Online Bagging processes data incrementally. As new data points are introduced, the algorithm continuously updates its models without the need to store or revisit prior data. The model is retrained multiple times with assigned weights for each incoming data point, thereby emulating the bootstrapping process. This characteristic renders it particularly suitable for real-time or streaming data scenarios. In our implementation of Online Bagging and the associated experiments, we utilized CapyMOA[4], a Python framework built upon the well-established MOA (Massive Online Analysis).

To accurately simulate the real-time nature of streaming data, we employed a train-test evaluation method specifically designed for stream data in these experiments. This approach sequentially utilizes each chunk of data for both training and testing purposes. Specifically, the model is trained on the i-th chunk and subsequently tested on the $i+1$-th chunk, allowing for progressive learning from the stream. This procedure is repeated for all data chunks, resulting in $N-1$ training and testing cycles for a total of N chunks. This methodology ensures that the model is assessed on unseen data, mirroring real-world scenarios where predictions are made based on future data after prior learning.

It is important to note that the features ‘id_student‘, ‘date‘, and ‘date plus‘ were excluded from the computations to prevent bias in the results. Additionally, the ‘feature score‘ was removed due to the presence of numerous missing values resulting from incomplete assessments.

Standard classification metrics were employed to evaluate the model’s performance after processing each chunk. These metrics include accuracy, precision, recall, and F1-score, providing a comprehensive assessment of the classifier’s performance as it sequentially processes data over time.

Table 2 presents the average classification performance along with the standard deviation across all chunks for each of the three experimental settings (months, trimesters, and semesters), reported separately for the two classes: Fail and Pass.

Given the significant class imbalance present in the dataset, accuracy as a metric is deemed unreliable, as it tends to be biased toward the majority class. Consequently, the following analysis emphasizes the F1-score, precision, and recall for both classes. The results indicate that the model encounters challenges in accurately identifying the Fail class, consistently recording low values across all three experimental settings. In contrast, the Pass class demonstrates high

[4]CapyMOA: https://capymoa.org/

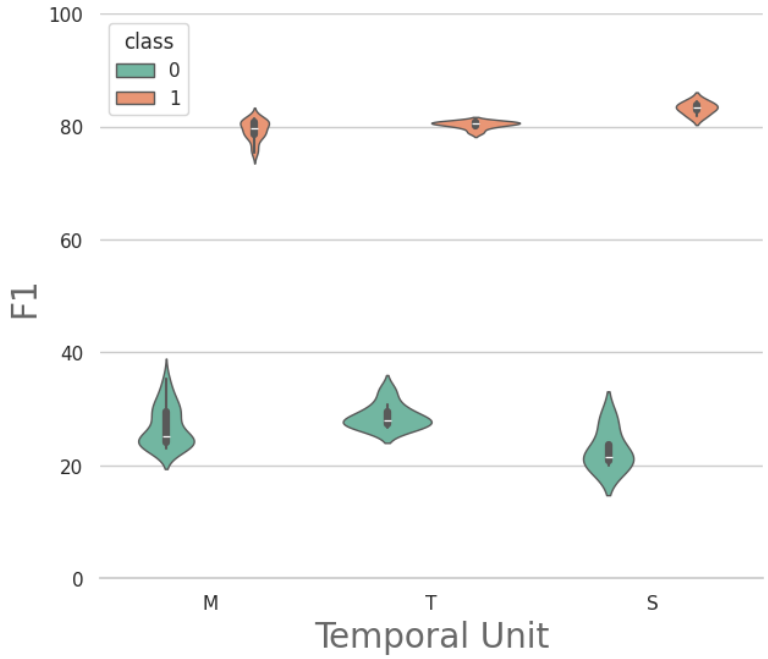


Figure 2: Violin Plots representing the distribution of the F1 values over the chunks in different periods: Months (M), Trimesters (T), and Semesters(S), obtained with the Online Bagging stream classifier.

scores for all three metrics. Notably, when employing the "semester" temporal unit, the recall for the Pass class reached 93.63%, signifying that nearly all samples labeled as Pass were correctly identified by the model.

A similar trend is evident across all three experimental settings: the models exhibit proficiency in recognizing the majority class while struggling with the minority class. Although the optimal results were obtained with semester-based data—likely due to the more comprehensive information contained within each semester—the performance associated with monthly and trimester data was only marginally lower. This observation suggests that the model effectively learned the underlying data structure and utilized it competently for predictions.

Nonetheless, the recall values for the Fail class remain notably low, particularly with semester-based data, where it was recorded at 14.31% with considerable variance. There is, however, a slight improvement in recall for the Fail class when using monthly and trimester data, achieving values of 20.92% and 22.88%, respectively.

To further analyze the average performance across all chunks in each experimental setting, Figure 2 presents a Violin Plot of the F1 scores for the two classes: 0 (Fail) and 1 (Pass). Violin plots effectively illustrate the data distribution across different categories by integrating features of both box plots and density plots. The central dot in the plot represents the median, while the thick bar indicates the interquartile range. The shapes extending from either side of the bar illustrate the data distribution, with the width signifying data density. Wider sections of the violin indicate higher data concentration, whereas narrower sections reflect sparser areas. Violin plots facilitate comparisons of central tendency and data distribution across multiple

groups.

Figure 2 corroborates the observations regarding the average F1 values for the Pass and Fail classes across the three temporal settings. This figure also elucidates the distribution of F1 scores obtained from different data segments within each experimental scenario. Elongated violins indicate high variability in the data, while wider and flatter violins suggest low variability, signifying that the model is robust to changes in the data. It is important to consider that the number of data segments varies across the three scenarios (30 for monthly data, 10 for trimester data, and 5 for semester data), which may influence the interpretation of the results.

Furthermore, as illustrated in Figure 1, while the Pass class consistently represents the majority in every segment across all time periods, the monthly dataset exhibits several instances of data drift. This drift significantly reduces the representation of the minority Fail class, thereby complicating accurate predictions (e.g., in segments X and Y). Although these drifts are still present in the trimester data, they occur to a lesser extent and are absent in the semester dataset. The more frequent drifts observed in shorter time intervals elucidate the more elongated shape of the violins for these periods.

Interestingly, the variability for the Pass class is minimized, with F1 scores deviating only slightly from the average despite some variance. In contrast, the Fail class exhibits considerably higher variability. Notably, in the semester-based experiments, the elongated violin for the Fail class indicates that, despite the larger data size compared to shorter periods, the data remains insufficiently informative for constructing an accurate predictive model for this minority class.

In summary, the new dataset proves valuable for predicting student outcomes. However, it is crucial to ensure that the data is balanced or that the differences between the two classes are minimized to achieve acceptable classification performance for both the Fail and Pass categories. Moreover, the results indicate that the various temporal periods provide adequate information for the classifier, thereby enabling continuous monitoring of student activities. This capability allows for timely interventions when a student's performance declines, facilitating support before issues escalate.

4. Conclusions and future work

Recent studies have indicated that stream algorithms can effectively monitor student learning within Virtual Learning Environments (VLEs) when incorporating temporal information. However, a significant limitation exists due to the lack of student data containing such temporal information. To address this issue, this study introduces a novel dataset derived from the Open University Learning Analytics Dataset (OULAD). We manipulated the timing information within OULAD to establish a unique temporal framework for student interactions with the platform.

To evaluate the utility of this dataset for predicting student outcomes in a streaming context, we conducted experiments using an online stream algorithm known as Online Bagging, implemented via the Python library CapyMOA.

Three experimental settings were designed, where data subsets were grouped by different temporal units: month, trimester, and semester. The results indicated that the classifier struggled to recognize the Fail class, which is underrepresented compared to the Pass class. However, the Pass class was correctly identified, achieving F1 scores of approximately 80% across all three

settings. Despite numerous fluctuations in the data—variability that increased as the temporal units decreased—the classifier demonstrated robustness in recognizing the Pass class. This finding highlights the potential of stream algorithms to effectively monitor student progress over time.

Research has shown that the choice of temporal units can significantly influence the outcomes of classification models aimed at predicting student performance. Our study revealed that even the smallest temporal unit of 30 days (equivalent to a month) yielded results comparable to those obtained with larger units. This suggests that data from a single month is sufficient for capturing learning patterns that are instrumental in predicting student performance, particularly when utilizing adaptive and incremental stream algorithms.

Nevertheless, this work represents an initial exploration of the new dataset. Future research will further investigate the application of stream algorithms within the educational domain, with a focus on enhancing their effectiveness and evaluating their applicability in real-time scenarios. This will involve analyzing data on a daily or weekly basis to promptly identify students at risk of failure and facilitate early intervention. Additionally, we aim to explore methods for enhancing the interpretability and reliability of results derived from streaming data. Finally, we plan to integrate resampling techniques into the streaming data context to address the challenges identified in this study, particularly those related to class imbalances and overall classification performance.

5. Acknowledgments

G. Castellano and G. Zaza acknowledge funding support from the FAIR - Future AI Research (PE00000013) project, Spoke 6 - Symbiotic AI (CUP H97G22000210007), under the NRRP MUR program funded by NextGenerationEU. All authors are members of the INdAM GNCS research group.

References

[1] L. Chen, P. Chen, Z. Lin, Artificial intelligence in education: A review, Ieee Access 8 (2020) 75264–75278.

[2] R. Alfredo, V. Echeverria, Y. Jin, L. Yan, Z. Swiecki, D. Gašević, R. Martinez-Maldonado, Human-centred learning analytics and ai in education: A systematic literature review, Computers and Education: Artificial Intelligence (2024) 100215.

[3] W. Chango, J. A. Lara, R. Cerezo, C. Romero, A review on data fusion in multimodal learning analytics and educational data mining, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 12 (2022) e1458.

[4] C. Romero, S. Ventura, Educational data mining and learning analytics: An updated survey, Wiley interdisciplinary reviews: Data mining and knowledge discovery 10 (2020) e1355.

[5] G. L. Bosco, G. Pilato, D. Schicchi, Deepeva: a deep neural network architecture for assessing sentence complexity in italian and english languages, Array 12 (2021) 100097.

[6] R. Avogadro, F. D'Adda, M. Cremaschi, Feature/vector entity retrieval and disambiguation

techniques to create a supervised and unsupervised semantic table interpretation approach, Knowledge-Based Systems 304 (2024) 112447.
[7] P. V. de Campos Souza, E. Lughofer, A. J. Guimaraes, Regularized neuro-fuzzy ai model to aid score management in online distance learning forums, in: 2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, 2021, pp. 1–8.
[8] D. Schicchi, D. Taibi, Ai-driven inclusion: Exploring automatic text simplification and complexity evaluation for enhanced educational accessibility, in: International Conference on Higher Education Learning Methodologies and Technologies Online, Springer, 2023, pp. 359–371.
[9] G. Abdelrahman, Q. Wang, B. Nunes, Knowledge tracing: A survey, ACM Computing Surveys 55 (2023) 1–37.
[10] G. Casalino, L. Grilli, P. Limone, D. Santoro, D. Schicchi, Deep learning for knowledge tracing in learning analytics: An overview, in: CEUR Workshop Proceedings, volume 2817, 2021.
[11] G. Casalino, G. Castellano, A. Mannavola, G. Vessio, Educational stream data analysis: a case study, in: 2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON), IEEE, 2020, pp. 232–237.
[12] D. Leite, I. Škrjanc, F. Gomide, An overview on evolving systems and learning from stream data, Evolving systems 11 (2020) 181–198.
[13] G. Casalino, P. Ducange, M. Fazzolari, R. Pecori, Incremental and interpretable learning analytics through fuzzy hoeffding decision trees, in: International Workshop on Higher Education Learning Methodologies and Technologies Online, Springer, 2022, pp. 674–690.
[14] G. Casalino, G. Castellano, D. Di Mitri, K. Kaczmarek-Majer, G. Zaza, A human-centric approach to explain evolving data: A case study on education, in: 2024 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS), IEEE, 2024, pp. 1–8.
[15] J. Kuzilek, M. Hlosta, Z. Zdrahal, Open university learning analytics dataset, Scientific data 4 (2017) 1–8.
[16] G. Casalino, G. Castellano, G. Zaza, A new dataset for streaming learning analytics, Zenodo, 2024. URL: https://doi.org/10.5281/zenodo.14003233, [Data set].
[17] N. Oza, S. Russell, Online bagging and boosting, in 'in artificial intelligence and statistics 2001', Morgan Kaufmann, str 105 (2001) 14.