Predicting At-Risk Students Using Weekly Activities and Assessments

https://doi.org/10.3991/ijet.v17i19.31349

Moohanad Jawthari[1(✉)], Veronika Stoffa[1,2]
[1]Department of Media and Educational Informatics, Faculty of Informatics, Eötvös Loránd University, Budapest, Hungary
[2]Department of Mathematics and Computer Science, Faculty of Education, Trnava University, Trnava, Slovakia
mjawthari@inf.elte.hu

Abstract—Although Online learning has been so popular especially during epidemic crisis, it has a drawback of high dropouts and low completion rates. Institutes search for ways to support their students learning and increase completion rates. Institutes will be able to predict students' performances and make interventions on time if they have some analytical strategy. Yet, efficient prediction and proactive intervention depends on using meaningful, reliable, and accurate data. Institutes different tools like Virtual Learning Environment (VLE) for teaching and content delivery. These tools provide large databases that are useful to improve prediction of students' performance research. In this study, an Open University course VLE data is analyzed to investigate if weekly engagement alone, integrated with assessments scores, and accumulated previous assessments up to a certain week data lead to accurate student performance prediction. Importance of VLE data is highlighted here, which sheds light on students' haviour and leads to developing models that can predict student's outcome accurately. Predictive models, Learning Analytics, Educational data mining, Classification.

Keywords—at-risk prediction, predictive models, learning analytics, educational data mining

1 Introduction

Low course enrollments and even lower course completion rates are posing problems for higher education institutions. Low dropout rates are quickly becoming a concern, and universities are looking for approaches to boost the retention rate of their students. To emphasize those challenges, OECD reported that only 31% of students completed a 4-year degree program in Australia, just 49% of students completed such program in US, and 71% was the completion rate for UK [1]. Lower retention rates risk a university financial stability in long term. Hence, universities concentrate on strategies which ensure students achievements and can deliver aggressive efforts to aid learners in their academic endeavors. Detecting at-risk students earlier and making on-time interventions have been effective in improving students' retention. Yet, proa tive action and helpful intervention needs an accurate, reliable, and meaningful data [2]–[4].

Higher education institutions can make timely interventions to improve students' performance by having some way of predicting students' performance. These institutions have benefited from the extensive usage of tools like Student Management Systems (SMS), Learning Management Systems (LMS), and Virtual Learning Environment (VLE) in offering smooth online communication, delivering learning and teaching resources, designing interactive learning activities, and managing academic assessment. In addition, those tools offer educational institutions demographics, student academic records, and log files datasets. These logs are based on students' interactions with the LMS, and they have provided us with new research avenues to help improving students' performance [4]–[5]. There have been numerous success stories about how data gathered from the tools that employed student data has helped in improving overall retention rates [6]. Georgia State University, for example, employed predictive analytics to boost graduation rates from 32% in 2003 to 54% in 2014 [7]. Moreover, at-risk students at Purdue University in USA were predicted as early as the second week, so these students got additional support and that resulted in improving their academic performance [8].

In this study, we will present comparative studies of models that have been used to predict students final course outcome based on assignments scores and engagement data from Virtual Learning Environment (VLE). This paper presents some preliminary results of an analytic strategy based on accumulating previous assessments in addition to VLE data. It also

2 Statement of the issue

Virtual Learning Environment is commonly utilized in higher education institutions to provide course contents. VLE, on the other hand, is capable of much more than merely delivering learning materials. VLEs supports teaching methods by providing an environment that supports online interactions with learning materials as well as improved communication between students and teachers. Activities carried out during the course, can be used to discover student behavior that can be used to predict expected future grades. Effective prediction, on the other hand, is dependent on the dataset; the larger the data set, the more accurate the model will be. Despite the fact that VLE provides a large number of learning activities, VLE tools are used in a variety of ways. For example, some institutes utilize VLE tasks heavily, while others just use it for communication. The extensive usage of VLE activities increases students' digital traces, which are extremely valuable in prediction. In this study, we have analyzed the VLE data of a distance learning course using two analytics strategies and have shown when the VLE data was integrated with assignment scores helped improve the prediction accuracy in first approach. Second approach has also shown accumulating previous weeks assessments in addition to VLE data led to a noticeable improvement in models predictions accuracy. Following are the research questions that are addressed in this paper.

Research Question 1: What is the earliest possible time span (week) to generate a robust prediction model that forecasts at-risk students of failing the course using assessments scores and engagement data?

Research Question 2: Does accumulating assessments from prior weeks beside engagement data generate a robust prediction model?

3 Dataset understanding and data preparation

3.1 Data understanding

Data for this study is obtained from Open University (OU) courses, one of the largest distance-base universities. The goal of developing Open University Learning Analytics dataset (OULAD) was to support the research in the area of learning analytics and educational data mining [9]. This dataset is unique as it contains demographic data together with aggregated clickstream data of students' interactions in the Virtual Learning Environment (VLE) and is freely available. The dataset includes information about 22 courses delivered between 2013–2014, 32,593 students, their assessment results, and their interactions logs with the VLE represented by daily summaries of student clicks (10,655,280 entries). The courses come from two disciplines: "Social sciences" and "Science, Technology, Engineering and Mathematics". In OU, courses are called models, which can be presented many times during the year. To differentiate between model's various presentations, the year and the starting month are used to name the module. For instance, presentations starting in January ends with A, in February with B and so on. Thus, '2013J' means that the presentation started in October 2013.

Figure 1 shows the database schema. The course table includes data about modules and their presentations. The registration table contains information about learners such as timestamps and course enrollment dates for the course. The assessment table contains information related to assessments such as assessment id, assessment type, and cut-off date for submission. Three types of assessments are available: tutor marked assessment, computer marked assessment, and the final exam. The learner-assessment table refers to the assessment results for each learner.

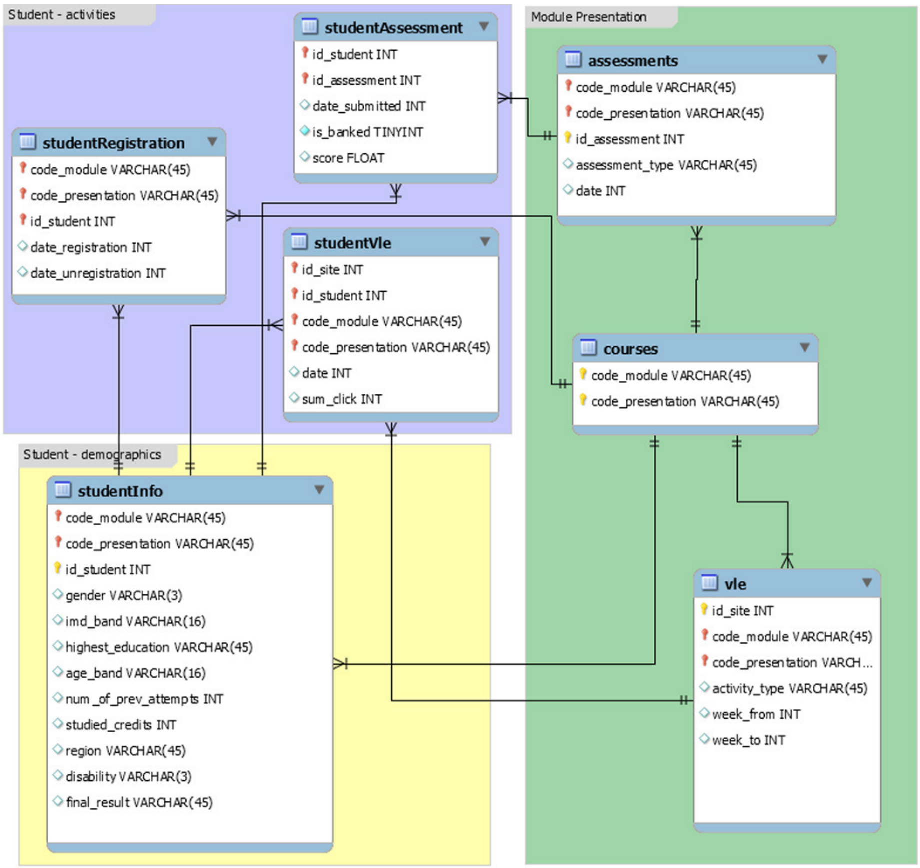

Fig. 1. OU schema [9]

3.2 Dataset and data preparation

This study explores data from the module called "DDD" and "2013B" presentation. This module belongs to "Science, Technology, Engineering and Mathematics" category. There were 1303 students enrolled in the module. Each enrolment is associated with a log of students activities which include watching lecture videos, working on course's problems, submitting assessments, accessing course's modules, discussing in forums and so on. The module has 536,837 VLE interactions with 4,903 learning activities, and 173,912 submitted assessments. Its duration is 240 days, i.e., 34 weeks. The reason to select the module is that it has high rates of fails and withdrawals; 432 learners dropped out and 361 failed, and large number of enrollments, 1303.

The course was arranged in a weekly structure, which meant that students were supplied with weekly learning materials, and they engaged with the new material and completed required tasks on a weekly basis. Logs were provided as daily total clicks;

however, to match the logs to the course structure we had aggregated the events into weekly format.

To use the data with machine learning models, relational databases tables should be mapped to tabular data. All the clicks recorded before the module starting day, 0, were filtered out. Formula 1 was applied to transform date feature in StudentVle and assessments tables from daily basis to weekly basis. StudentVle and vle tables were joined to obtain activity types for students clicks. Figure 2 illustrates the steps taken to produce the final behavioral table. A week engagement data was formed using formula 2. For example, $w_2 = w_0 + w_1 + w_2$. Figure 3 shows the engagement level for each students group based on their results in the course. As can be seen, students who engaged more passed the course.

$$date = \lfloor date / 7 - 1 \rfloor + 1 \quad (1)$$

$$w_n = \sum_{i=0}^{n} W^i \quad (2)$$

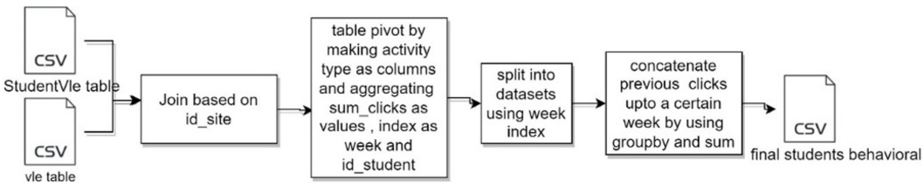

Fig. 2. Engagement data linking

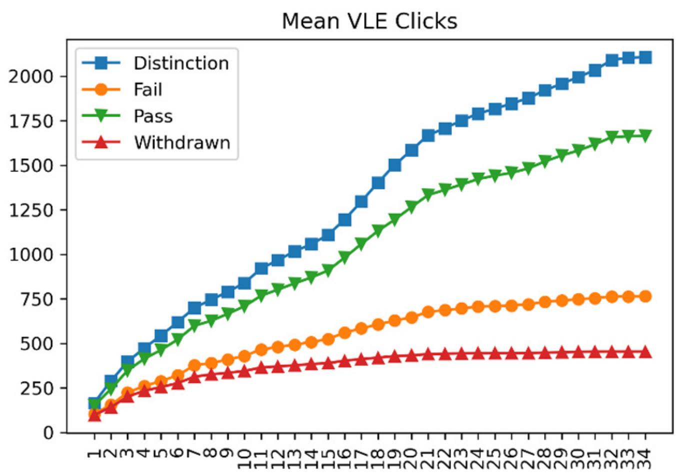

Fig. 3. Average VLE clicks during the course weeks based on final result (target variable)

To obtain assessemnts in the form like behavioral table, student Assessment and assessmnets tables were joined based on id_assessmnet. The scores of assessments are provided as out of 100 and have different weights. We applied formula 1 to create an aggregated score feature in the weeks that have assessments. n denotes the number of the assessments in the week, and s_i denoted the ith assessment score and w_i represents its wight.

$$w_{assessmnets-score} = \sum_{i=1}^{n} \frac{s_i * w_i}{100} \quad (3)$$

There are thirteen assessments (CMA and TMA) beside Exam. First assessments, both CMA and TMA were due in the third week of the module. Other assessments were due in weeks 7, 11, 16, 21, 24, 28, and 29. Exam was due in the last day of the course. Once the results of assessments scores become available, they are included in the VLE engagement data to train the model. Students have been divided in different groups based on their final grade in the course. Figure 4 shows average scores in assessment; X axis shows the assessment type and when it was available.

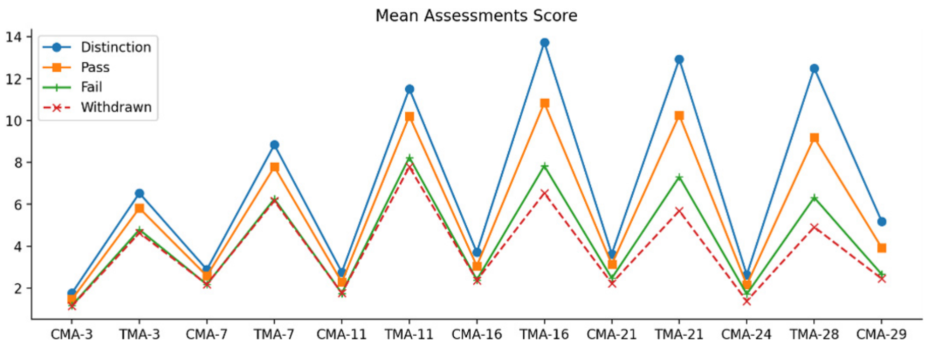

Fig. 4. Average assessments scores based on final result (target variable)

Table 1 illustrate the difference between the two approaches, table shows the 4th week dataset of approach 2 which contains w3 that is the assessments' score of week 3. On the other hand, approach 1 does not contain any assessments score during that week and did not accumulate the assessments scores of previous weeks (i.e., from week 3).

Table 1. Sample of final dataset after mapping

Id_Student	Externalquiz	Forumng	Glossary	Homepage	w3	…	Final_Result
40419	1	56	0	22	7.04		1
41060	2	5	0	14	7.915		1
43284	4	35	0	68	7.385		1
45664	2	91	0	82	5.89		0

4 Methodology

4.1 Experimental design

This part overviews the methods which are used in this study i.e., classification. Classification is a task for data analysis that uses data to train a model utilized to identify and assigns categories to the data. Such models are called classifiers that predict the class labels. The process of classification includes two steps: learning step in which the classifier is constructed using training dataset and classification step in which the model is applied to assign labels to unseen data as illustrated in Figure 5. Training data includes identifying attributes set with their values, where one of the distinguished attributes is knows as label or class. A classification model learns from training data and is used to predict the data which was not part of the training data. Model's performance is measured as the number of correctly classified test data samples and is known as accuracy. In addition, kappa coefficient, F1-score, AUC (Area under curve) are other measures.

In this study, the following classification models are used due to their popularity for such problems: Random Forest (RF), Nave Bayes (NB), Logistic Regression (LR). Scikit-learn is the implementation source of these models; the models were trained using default parameters.

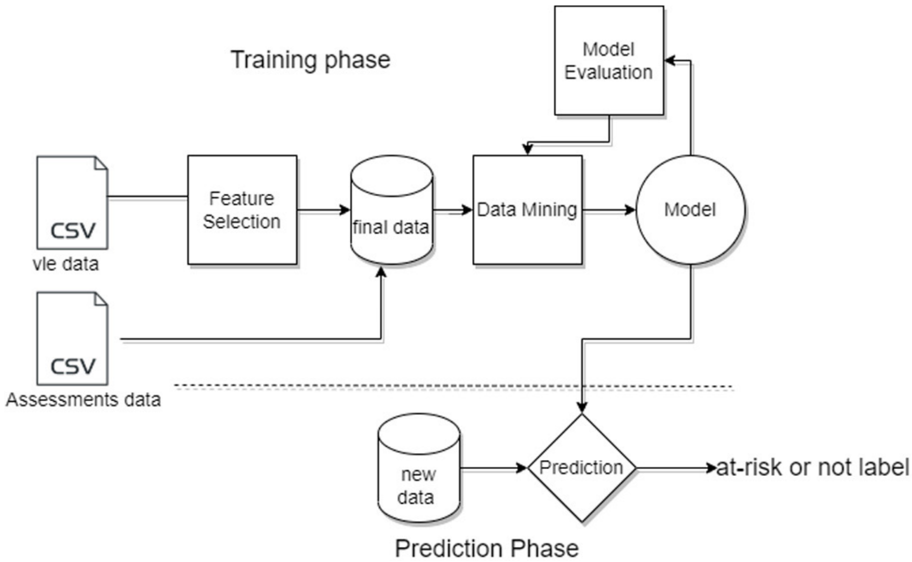

Fig. 5. At-risk prediction model

4.2 Evaluation metrics

To determine the efficiency of predictive models, F-Measure is used [10]. F-measure is commonly used with binary classification, especially in cases where dataset is imbalanced [11]. F-Measure is a test's accuracy measure that is calculated by considering precision P and recall R of the test. It is defined as the weighted harmonic mean of precision and recall shown in following equation.

$$\text{F-Measure} = 2.\frac{P.R}{P+R}$$

The maximum potential value of an F-measure is 1 and results from perfect precision and recall while 0 is the lowest possible value. Precision determines what ratio of positive identifications was correct. Recall determines what ratio of actual positives which were correctly identified. They are defined as following:

$$P = \frac{TP}{TP+FP} \text{ and } P = \frac{TP}{TP+FP}$$

where: TP (True Positives) = # students with label at-risk and were predicted as at-risk

TP (False Positives) = # students with label not at-risk and were predicted as at-risk

FN (False Negatives) = # students with label not at-risk and were predicted as the label at-risk.

4.3 Tools

Python module scikit-learn, [13], is for conducting experiments. Scikit-learn combines different machine learning algorithms to be used for supervised and unsupervised problem: classification, regression, and clustering. It is open-source software written in Python with a user friendly and task-oriented interface. Furthermore, matplotlib module was used for plotting graphs we used python [13].

4.4 Training process

Models were generated using three classifiers: RF, LR, and NB by using 10-fold cross validation technique. F-measure and Accuracy metrics are used evaluate ans compare models' prediction. These models are used to predict students' final out-come in the course and to classify it in two classes: 0 and 1, where 0 represents students who are not at-risk and 1 represents at-risk students. Prediction was performed based on the count of VLE activities in each week and assignment scores were added whenever they were available. Also, it was performed based on adding previous weeks assessments scores in a certain week.

5 Results and discussion

The results of the conducted experiments are discussed in this section.

5.1 Research question 1

Prediction was performed using the following machine learning techniques: NB, RF, and LR. The prediction aim was to predict final outcomes of students in the course in two classes: at-risk or not-at-risk. In this part, prediction was done on weekly bases; only VLE engagement data and assessments scores (if available) used in each week. The module has thirteen assessments and one final exam which was due in the last week of the semester. We used F-measure and accuracy to compare the results of models. As can be seen in Figure 6, F-measure scores increased in the weeks that had assessments available. For example, week 3 score was 0.768 which was achieved by LR model. This means assessments scores are discriminative. Figure 7 shows the accuracy results for models' performance. Accuracy is also improved in weeks that have assessments. LR accuracy in week 3 was 0.749 and outperformed other models during that week. Other weeks measures for both F-measure and accuracy are listed in Table 2. In week 2 (only VLE data), all models had close scores that are acceptable except RF. This means VLE alone can be used for the task of predicting at-risk students.

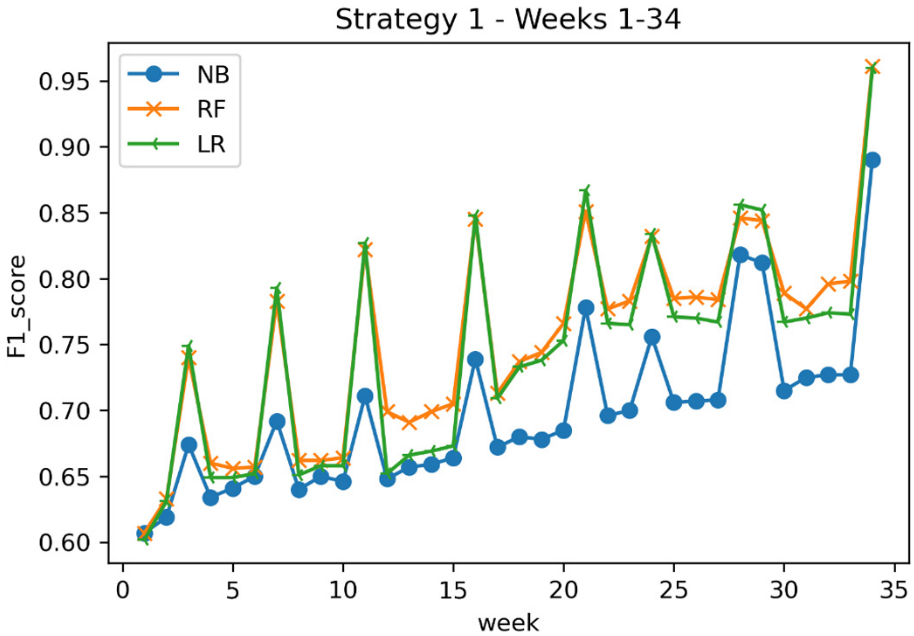

Fig. 6. F-measure results along weeks for approach 1

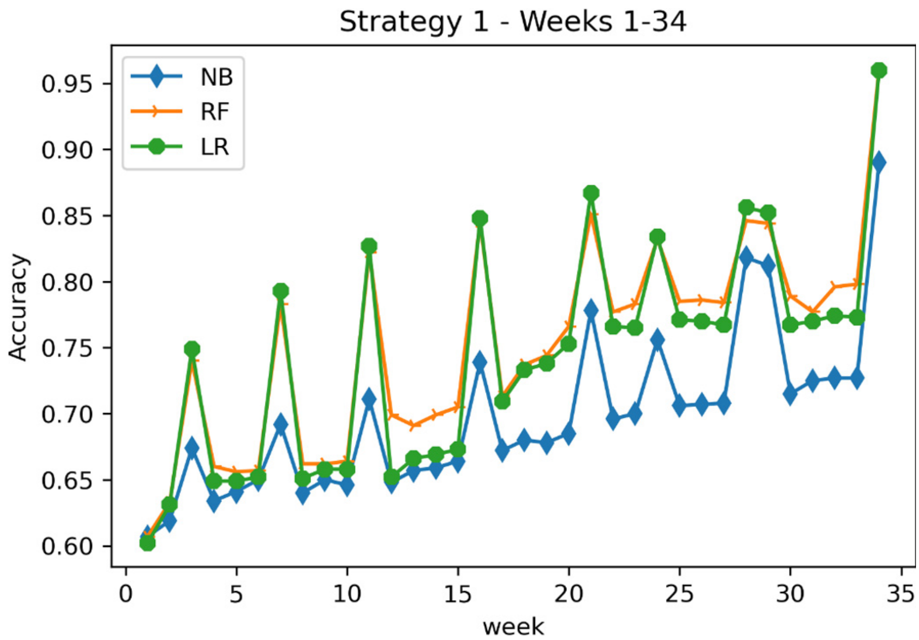

Fig. 7. Accuracy results along weeks for approach 1

Table 2. Models scores for approach 1 for all weeks

Week	F-Measure			Accuracy		
	NB	RF	LR	NB	RF	LR
1	0.696	0.636	0.694	0.607	0.607	0.602
2	0.712	0.677	0.717	0.619	0.633	0.631
3	0.745	0.766	0.768	0.674	0.74	0.749
4	0.723	0.696	0.726	0.634	0.66	0.649
5	0.726	0.689	0.725	0.641	0.656	0.649
6	0.729	0.692	0.724	0.65	0.657	0.652
7	0.749	0.805	0.808	0.692	0.783	0.793
8	0.721	0.691	0.723	0.64	0.662	0.651
9	0.73	0.69	0.728	0.65	0.662	0.658
10	0.726	0.694	0.725	0.646	0.664	0.658
11	0.769	0.838	0.84	0.711	0.822	0.827
12	0.73	0.72	0.721	0.648	0.699	0.652
13	0.737	0.714	0.732	0.657	0.691	0.666
14	0.739	0.722	0.736	0.659	0.699	0.669
15	0.743	0.725	0.738	0.664	0.705	0.673

(Continued)

Table 2. Models scores for approach 1 for all weeks *(Continued)*

Week	F-Measure			Accuracy		
	NB	RF	LR	NB	RF	LR
16	0.791	0.86	0.864	0.739	0.845	0.848
17	0.749	0.734	0.763	0.672	0.713	0.709
18	0.755	0.76	0.782	0.68	0.737	0.733
19	0.754	0.769	0.785	0.678	0.744	0.738
20	0.758	0.787	0.795	0.685	0.766	0.753
21	0.817	0.867	0.881	0.778	0.851	0.867
22	0.766	0.794	0.805	0.696	0.777	0.766
23	0.769	0.802	0.804	0.7	0.783	0.765
24	0.805	0.848	0.853	0.756	0.832	0.834
25	0.774	0.802	0.808	0.706	0.785	0.771
26	0.775	0.803	0.807	0.707	0.786	0.77
27	0.775	0.801	0.804	0.708	0.784	0.767
28	0.847	0.862	0.878	0.818	0.846	0.856
29	0.843	0.859	0.873	0.812	0.844	0.852
30	0.779	0.806	0.803	0.715	0.789	0.767
31	0.787	0.792	0.805	0.725	0.777	0.77
32	0.787	0.811	0.809	0.727	0.796	0.774
33	0.788	0.813	0.809	0.727	0.798	0.773
34	0.903	0.965	0.964	0.89	0.961	0.96

5.2 Research question 2

In this section, we discuss the experiments results and answer the second research question. In a similar way to the previous section, prediction was done on weekly bases; the first two weeks has only VLE, third week has the first assessments. The main difference is that 4,5, and 6 weeks included assessments of week 3. Figure 8 shows in week 4, f-score improved by 5%. The most noticeable prediction score belongs to RF model which is 0.78 compared to 0.688 in strategy 1. Predictions models improved in weeks 5 and 6 too. Week 7 data included week 3 assessments in addition to the available assessments during that week. Also, figure illustrates how the accuracy metric results increased in this approach. Table 3 shows the results of remaining weeks. As mentioned above, the improvement in scores means assessments are discriminative. This strategy makes at-risk prediction robust starting from the first time the assessments become available. Table 3 show that this method outperformed method 1 in terms of F-scores starting from week 4 until last week when the final exam was available.

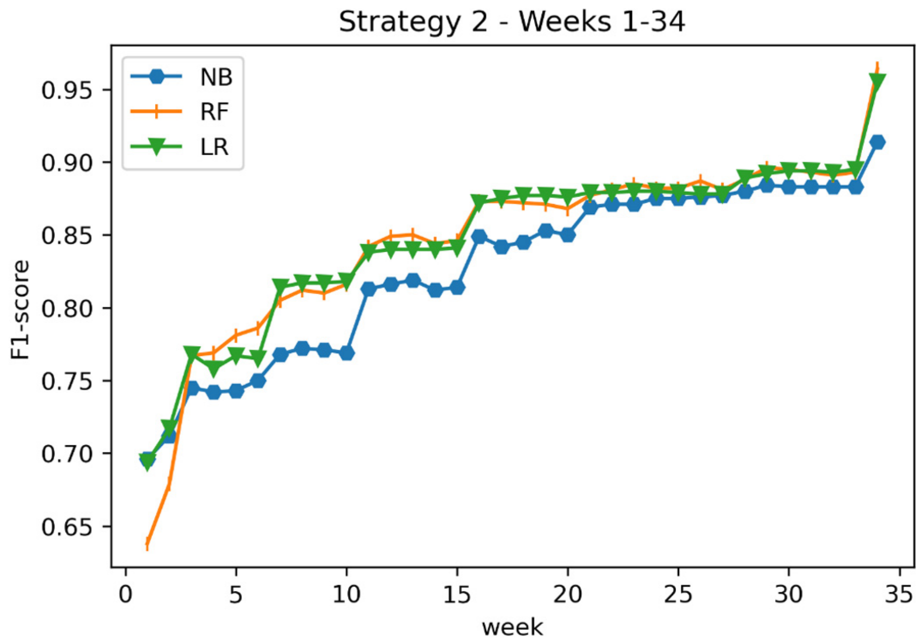

Fig. 8. Approach 2 F-measure results for all weeks

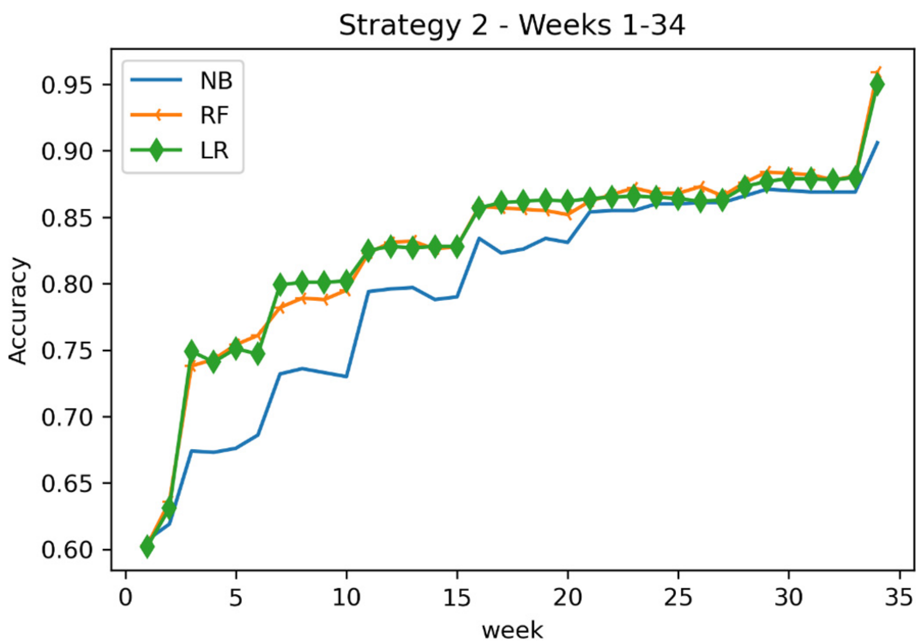

Fig. 9. Approach 2 accuracy results for all weeks

Table 3. Models scores for approach 2 during all weeks

Week	F-Measure			Accuracy		
	NB	RF	LR	NB	RF	LR
1	0.696	0.638	0.694	0.607	0.603	0.602
2	0.712	0.679	0.717	0.619	0.636	0.631
3	0.745	0.767	0.768	0.674	0.738	0.749
4	0.742	0.769	0.758	0.673	0.743	0.741
5	0.743	0.781	0.767	0.676	0.754	0.751
6	0.75	0.786	0.765	0.686	0.761	0.747
7	0.768	0.805	0.814	0.732	0.782	0.799
8	0.772	0.812	0.817	0.736	0.789	0.801
9	0.771	0.81	0.817	0.733	0.788	0.801
10	0.769	0.816	0.818	0.73	0.795	0.802
11	0.813	0.842	0.838	0.794	0.823	0.825
12	0.816	0.849	0.84	0.796	0.831	0.828
13	0.819	0.85	0.84	0.797	0.832	0.827
14	0.812	0.844	0.84	0.788	0.826	0.828
15	0.814	0.846	0.841	0.79	0.828	0.828
16	0.849	0.873	0.872	0.834	0.858	0.857
17	0.842	0.873	0.875	0.823	0.857	0.861
18	0.845	0.872	0.877	0.826	0.856	0.862
19	0.853	0.871	0.877	0.834	0.855	0.863
20	0.85	0.868	0.876	0.831	0.852	0.862
21	0.869	0.877	0.879	0.854	0.862	0.864
22	0.871	0.881	0.879	0.855	0.867	0.865
23	0.871	0.885	0.88	0.855	0.872	0.866
24	0.875	0.882	0.88	0.86	0.868	0.865
25	0.875	0.882	0.879	0.86	0.868	0.864
26	0.876	0.887	0.878	0.861	0.873	0.862
27	0.877	0.881	0.878	0.861	0.866	0.863
28	0.88	0.888	0.889	0.866	0.876	0.873
29	0.884	0.896	0.892	0.871	0.884	0.877
30	0.883	0.895	0.894	0.87	0.883	0.879
31	0.883	0.893	0.894	0.869	0.882	0.879
32	0.883	0.891	0.893	0.869	0.878	0.878
33	0.883	0.893	0.895	0.869	0.881	0.88
34	0.914	0.964	0.955	0.906	0.959	0.95

6 Conclusion

A comparative analysis of three predictive models namely Random Forest (RF), Naive Bayes (NB), was conducted in this paper. The goal was to predict at-risk students failing the courses by forecasting their outcome (at-risk or not at-risk). Prediction was done on weekly basis by using VLE data and assessments grades as first approach, and by accumulating assessments grades in the subsequent weeks as second approach. According to results, assignment scores are more discriminative than engagement data alone, and adding previous weeks assessments improved prediction scores. Nevertheless, by using VLE data alone, a prediction score was almost 70% in first week, and 71% in second week. Once assessments were available, the performance was boosted. The results are promising for virtual education where learners engage with learning activities; prediction performances depend on the behavioral data of students. Results also show the RF performance improved with accumulating previous assessments.

7 References

[1] OECD. Publishing, Education at a Glance 2014: OECD indicators, OECD, 2014.
[2] Scott, G., Shah, M., Grebennikov, L., & Singh, H. (2008). Improving student retention: A University of Western Sydney Case Study. Journal of Institutional Research, 14(1), 9–23.
[3] Tinto, V. (2012). Enhancing student success: Taking the classroom success seriously. Student Success, 3(1), 1. https://doi.org/10.5204/intjfyhe.v3i1.119
[4] Tinto, V., & Pusser, B. (2006). Moving from theory to action: Building a model of institutional action for student success. National Postsecondary Education Cooperative, 1, 51. https://doi.org/10.1007/978-90-481-8598-6_2
[5] Cavus, N., & Zabadi, T. (2014). A comparison of open source learning management systems. Procedia-Social and Behavioral Sciences, 143, 521–526. https://doi.org/10.1016/j.sbspro.2014.07.430
[6] Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), e1355. https://doi.org/10.1002/widm.1355
[7] Siemens, G. (2012, April). Learning analytics: Envisioning a research discipline and a domain of practice. In Proceedings of the 2nd international conference on learning analytics and knowledge, pp. 4–8. https://doi.org/10.1145/2330601.2330605
[8] Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittelmeier, J., … & Vuorikari, R. (2016). Research evidence on the use of learning analytics: Implications for education policy.
[9] Sclater, N., Peasgood, A., & Mullan, J. (2016). Learning analytics in higher education. London: Jisc. Accessed February, 8(2017), 176. https://doi.org/10.4324/9781315679563
[10] Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017). Open university learning analytics dataset. Scientific Data, 4(1), 1–8. https://doi.org/10.1038/sdata.2017.171
[11] Tan, P. N., Steinbach, M., & Kumar, V. (2016). Introduction to data mining. Pearson Education India.
[12] Olson, D. L., & Delen, D. (2008). Advanced data mining techniques. Springer Science & Business Media.
[13] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. The Journal of Machine Learning research, 12, 2825–2830.

8 Authors

Moohanad Jawthari, Department of Media and Educational Informatics, Faculty of Informatics, Eötvös Loránd University, Budapest, Hungary.

Veronika Stoffa, Department of Media and Educational Informatics, Faculty of Informatics, Eötvös Loránd University, Budapest, Hungary; Department of Mathematics and Computer Science, Faculty of Education, Trnava University, Trnava, Slovakia.

Article submitted 2022-03-31. Resubmitted 2022-06-13. Final acceptance 2022-06-13. Final version published as submitted by the authors.