Estimation of Success in Collaborative Learning based on Multimodal Learning Analytics Features

Daniel Spikol
Deparment of Computer Science
Malmö University, Sweden
Email: daniel.spikol@mah.se

Emanuele Ruffaldi & Lorenzo Landolfi
PERCRO
Scuola Superiore Sant'Anna, Italy
Email: e.ruffaldi@santannapisa.it

Mutlu Cukurova
UCL Knowledge Lab
University College London, United Kingdom
Email: M.Cukurova@ucl.ac.uk

Abstract—**Multimodal learning analytics provides researchers new tools and techniques to capture different types of data from complex learning activities in dynamic learning environments. This paper investigates high-fidelity synchronised multimodal recordings of small groups of learners interacting from diverse sensors that include computer vision, user generated content, and data from the learning objects (like physical computing components or laboratory equipment). We processed and extracted different aspects of the students' interactions to answer the following question: which features of student group work are good predictors of team success in open-ended tasks with physical computing? The answer to the question provides ways to automatically identify the students performance during the learning activities.**

Keywords—*Multimodal learning analytics; collaborative learning; practice-based learning*

I. INTRODUCTION

Over the last several years the field of learning analytics has grown rapidly in conjunction with Massive Open Online Courses (MOOCs) and other technology systems (mobile applications, student response systems) that have become part of the everyday educational landscape. These systems provide diverse types of data about learners' interactions in digital environments and allow new insights into education. Such systems highlight the importance of data in education that is of interest to diverse actors for utilising learning analytics for educational management and policy [1]. From a research perspective, the aim of learning analytics is to understand and optimise learning [2] however, the majority of learning takes place in face-to-face activities. In this paper, we focus on the potential of learning analytics on providing insights about practice-based learning, in which small groups of learners create unique solutions to open-ended problems.

Project-based learning activities are considered to have the potential to help educators achieve high tier institutional and policy goals such as developing 21st century skills in STEM subjects at scale. These activities, which often combine physical computing technologies with design principles in the context of computer science education, are increasingly popular in both secondary and post-secondary learning institutions [3]. To better accommodate learning in small groups, researchers use low-cost sensors, inexpensive computational power, and data from diverse sensors that include computer vision, audio, biometric, and data from the learning objects (like physical computing components or laboratory equipment) to collect data. The multimodal data from these sensors provides new opportunities for investigating learning activities in the real-world between small groups of learners working on tasks with physical objects [4], [3].

The aim of this paper is to investigate how Multimodal Learning Analytics (MMLA) can be used to support project-based learning from a specially designed worktable environment where small groups of students (three students) use new physical computing components to solve open-ended tasks. The ability to collect multimodal data from bodily movements, face tracking, affective sensors, log files from the hardware and software, user and research generated data provide opportunities to obtain useful features for understanding. Through the use of multimodal learning analytics platform that is part of the worktable, we collected diverse streams of data from learning activities. We processed and extracted multimodal interactions to answer the following question: *which features of student group work are good predictors of team success in open-ended tasks with physical computing?* Potential answers to this question would enable us to generate analytics to predict the success of groups based on multimodal features automatically. Such predictions are useful for both improving the awareness of students and practice of teachers to adjust their interventions.

II. BACKGROUND

Project-based learning activities that support learners' participation in open-ended tasks are one of the central teaching approaches in STEM education. These activities emphasise the engagement of learners in projects that are personally meaningful, support engagement in authentic inquiry activities and are collaborative [5]. However, the existing evidence about the impact of these approaches on learning is rare and one potential for this lack of evidence might be that the hands-on and open-ended nature of project-based learning creates challenges for tracking the learning processes. Moreover, often the value of such approaches is hard to measure with existing standardised measurement methods [4]. More recently, Multimodal learning analytics (MMLA) provide opportunities to help address these challenges and observe the evidence on expected learning outcomes of these complex pedagogies. This evidence and insights generated can be used to support students' learning processes in these approaches and improve the practitioners' practice.

Current research has focused on MMLA as a means to better understand the complexity of learning through the advances of high-frequency multimodal data capture, signal

processing, and machine learning techniques [6]. There is an emerging body of work with MMLA to capture small group work on project-based learning that has grown out of the work of Blikstein and Worsley [7], [8]. Blikstein [9] explored multimodal techniques for capturing code snap-shots to investigate students learning computer programming and video and gesture tracking for engineering tasks. This work offers an opportunity to capture different insights about learning in tasks in which students have the opportunity to generate unique artefacts like computer programs, robots, and collaborate in small groups to solve open-ended tasks. Worsley [10] presented different approaches for data integration and fusion and how these can have a significant impact on the relation of research and learning theories. These approaches provided the means for other researchers to begin to explore MMLA with small groups of students across different subjects. Grover and colleagues [11] have explored how to develop test computational models of social learning in collaborative problem-solving environments. Their approach has been to classify the quality of collaboration from body movement and gestures of pair programmers working together. Prieto and colleagues [12] and Martinez-Maldonado and colleagues [13] have focused their research efforts on how MMLA can support teaching actions and orchestration in the classroom. Drawing from the literature we can observe that MMLA has a role to play to support education in project-based learning through providing new means for gathering insights for complex open-ended tasks [4].

III. SYSTEM AND CONTEXT

The work discussed in this paper is based on a European project "Practice-based Experiential Learning Analytics Research And Support" (PELARS)[1]. One of the aims of the project was to develop learning analytics tools for hands-on, open-ended STEM learning activities using physical computing. The learning contexts investigated are STEM education in high school; engineering and interaction design at universities. The current system includes customised furniture with an integrated Learning Analytics System (LAS) including tracking hands, faces and other objects, meanwhile the Arduino platform with a visual web-based Integrated Development Environment (IDE) captures interaction information of physical computing. The learners and observers use mobile devices to capture multimedia data (text, images, and video) to self-document the learning activities. Overall, the PELARS project developed an intelligent system for collecting activity data for diverse learning analytics (with data-mining, reasoning, and visualisations) and active user-generated material and digital content (mobile tools) from practice-based activities (physical computing platform)[14].

LAS collects multimodal data from different sensors and inputs from the learners and researchers. The learning environment is designed to foster collaboration and includes an integrated screen and standing round table to allow learners to share and work together. The LAS collects data from both ambient (sensors) and live sources (human interaction). The ambient collection of data includes a computer vision system that uses both an RGB with depth, video with audio and video that collects data about how people interact around the
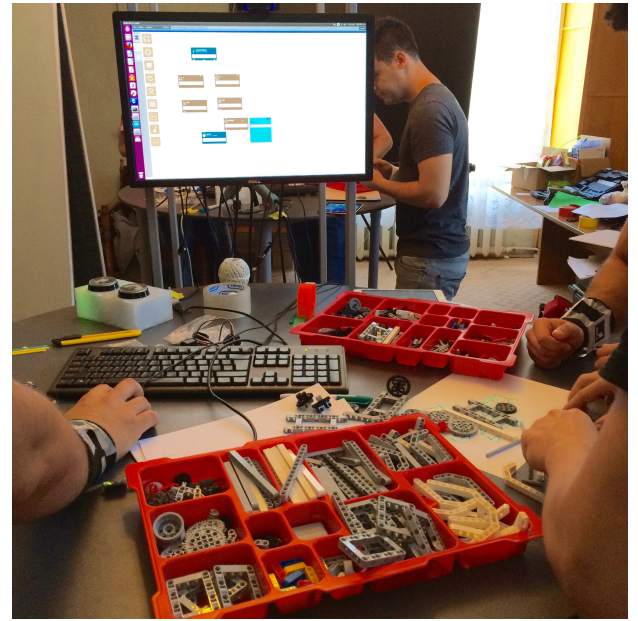
Fig. 1. University engineering students working in the PELARS environment.

workstation furniture. The specific implementation relies on client-server architecture where a cloud infrastructure collects the data processed by the local client, called collector. The communication is based on WebSockets and supports disconnection and offline modality taking into account the connectivity challenges of learning environments. Analytics and machine learning are performed in the cloud. The PELARS system in action in Figure 1.

IV. MATERIALS AND METHODS

The automatic approach discussed in this paper is performed over a data set acquired with students with the PELARS platform. In this section the data acquisition processes are discussed, with the analysis performed based with machine learning classification.

A. Datasets

The data analysed in this paper is from 3 sessions for each of the 6 groups made of 3 students studying engineering at a European University (average age 20 years old with 17 men and 1 woman), for a total of 18 sessions. Each student group used the system over 3 days completing one open-ended tasks in each session. The students were introduced to the system and then their first task was to prototype an interactive toy. The second task was the prototyping a colour sorter machine, and the third task targeted an autonomous automobile. Each of the tasks introduced more complex concept to be solved with respect to the previous ones. Students were asked to perform an initial phase of planning, followed by execution and finally a documentation phase. No specific instructions about the timing of these phases were given to students. During the session, the observer has split the session into these three phases based on personal judgment of the progress via the mobile tool.

To grade the sessions, a scoring scheme was developed that combined different approaches for collaborative problem solving (CPS) in small groups. We started with the seminal work done with engineering students [15] that was initially adopted by [10] for multimodal learning analytics. From these initial frameworks, we began to develop a framework for CPS [16] that we could apply for the PELARS context. We used a version of our CPS framework with the mobile system with an agreed set of codes for on-fly observations to initially grading of the students' projects. From the initial score of the students' work, the team of researchers reviewed the students work collected in the LAS which included snapshots of the students' plan, video of solutions, and learners text input. The 18 session were graded with these criteria, where 50% of the grade was the expert's opinion, 25% was how the students planned and delivered the artefact, and the remaining 25% was the students' own self-assessment. The resulting scores were categorised in three classes: poor, ok, and good equivalent to a Likert scale of 3 values.

Acquired Data For each sessions recorded, the LAS system collected data from the students comprising activity performed, user generated content (text and multimedia) and actions on the Arduino visual IDE. In particular, the following data was acquired.

Face Tracking By means of a frontal camera and the Viola-Jones algorithm the face of students was tracked, and then, thanks to camera calibration and assumptions about sizes, it was possible to estimate the distance of students from the camera. This means that the position of the face is computed in 3D coordinate with respect to a position on the table. Two metrics have been identified: the first is the count of Faces Looking at the Screen **FLS**, the second is the distance between the faces providing an indicator of Distance between Learners **DBL**. We hypothesise that the measure **DBL** may be a proxy of collaboration, since when **DBL** is small it is more like that the collaboration would occur among students.

Hand Tracking A top down camera monitored the motion of the hands of the students that were wearing fiducial markers that disambiguate each primary hand. Again, thanks to the calibration of the camera and the size of the markers the 3D position of the hands was obtained. The resulting metrics were the Distance between Hands **DBH** and the Hand Motion Speed **HMS**.

Arduino IDE The interface between the Visual Arduino IDE and the data collection system provided information about the types of physical and software blocks used in the project and their connections. In particular we counted the number of Active Blocks **IDEC**, the Variety of Hardware **IDEVHW** and Software Blocks used **IDEVSW** and the number of interconnections between blocks as a Measure of Complexity **IDEX** in students' programming during their practice-based activities.

Audio Level By means of the microphone included in one of the cameras and Fast Fourier Transformation (FFT) we computed the sound level during the sessions. The resulting feature was a value sampled at 2Hz called Audio Level **AUD**.

B. Preprocessing

From all these MMLA data points the data was collected at variable data rates (around 2Hz), yet it was not synchronised. For this reason, we needed a processing stage that aggregates indicators from the different variables in windows of same durations. The aggregation performed was based on counting for most of the variables, except for the distance/proximity functions for which we employed averaging. Taking into account the different lengths of the sessions, we employed zero padding for sessions that were too short. In our investigation, the tested window sizes were 10,20 and 30 minutes and there was also the case of one single window for the whole session.

C. Methods

The approach proposed in this paper is based on a supervised classification task that matches the observers' scores. The purpose of this approach is to identify the data features that can support different score classifications (poor, ok, good). Among the different families of classifiers, we opted for the parametric ones namely Naive Bayesian (NB), Logistic Regression (LR) and Support Vector Machines with linear (SVML) and Gaussian kernel (SVMR). We avoided the non-parametric ones (Nearest neighbours) or decision trees with the purpose of reducing the overfitting effect. In particular the Naive Bayesian is a simple classifier that employs a strong assumption about the feature, a condition that holds for most of the variables employed except for the ones related to the Arduino IDE. We decided not to use the ensemble of classifiers [17], as we would like to study the model behind these classifications as much as performing the classification itself.

We used cross-validation (k=4) for understanding the effect of different parameters such as window size and the inclusion of different phases. Due to the small sample size (18), we avoided the leave-one-out scheme. The data acquired from the PELARS LAS was exported and then processed in Python using the sklearn [18] toolkit that provides state-of-the-art machine learning techniques integrated with a common interface. The test of the classifiers was performed by varying the window size, the score (binary or original 3-level), the inclusion of the different phases (planning, building, and reflecting) and, most importantly, the effect of features described above (**FLS, DBL, DBH, HMS, IDEC, IDEVHW, IDEVSW, IDEX, AUD**)

V. RESULTS

The 18 session lasted between 33 and 75 minutes (median 63 min±13) for a total activity time of 17 hours and 10 minutes. The teams scored with different patterns along the three sessions and Figure 2 shows the outcomes.

The phases annotated by the observer were quite variable: planning (11min±10min), work (41min±16min) and reflection (4min±7min).

Below the results are being presented.

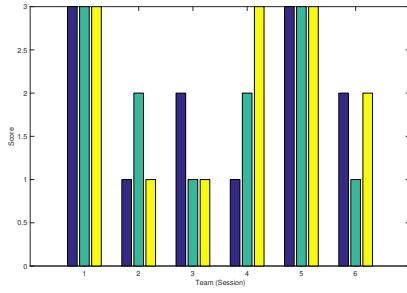Fig. 2. Scores of Teams across sessions

Fig. 3. Example of face distance in four sessions (top two are good plus OK sessions, bottom two are poor ones)

TABLE I. EFFECT OF PHASES IN THE INCLUSION OF THE CLASSIFIER. P=PLAN, W=WORK, R=REFLECT

	PWR	PW	W	WR
NB	0.8	0.8	0.6	0.75
SVML	0.6	0.75	0.75	0.8
SVMR	0.75	0.75	0.75	0.75
LR	0.6	0.75	0.5	0.6

A. Scoring

The three-level scoring (poor, ok, good) posed difficulties to the classification activity and we needed to move to a binary version in which we aggregated ok graded groups with good graded groups. For example, NB and SVM scores 0.8/0.75 with a window of 30min and binary classification, however this value decreases to 0.5 for a three-way classification. This is not ideal, however in order to achieve adequate results we took this binary approach. We see this as the first step towards further more detailed classifications. For the rest of the paper, we'll focus on this **binary** classification between poor graded groups vs "good and ok" graded groups.

B. Effect of Phase

Across the different conditions, the selection of the phases used to train a strong affect of the capacity to recognise the classifiers. For example, with a 30min window and binary classification, the exclusion of reflection (PW) phase in student activities, provided stronger results across the different classifiers, while the exclusion of both planning and reflection reduced the classification power. See Table I for the details.

In order to provide the most reliable results and use the strongest classification power, for the rest of the paper, we focus our results on planning and working stages of the student activities and excluding the reflecting stages.

C. Type of Classifiers

As can be seen in table 1, across the different tests the classifier, those behaved the most consistently were Naive Bayesian (NB) and the Support Vector Machines with linear kernel (SVML).

D. Effect of Features

Having established the window size (30 mins), grade classifications (poor vs. ok plus good scores), learning activity stages (planning and building), and the statistical methods we will use (NB and SVML), we now present the results of our analysis on the effect of the learning analytics features. We start from the f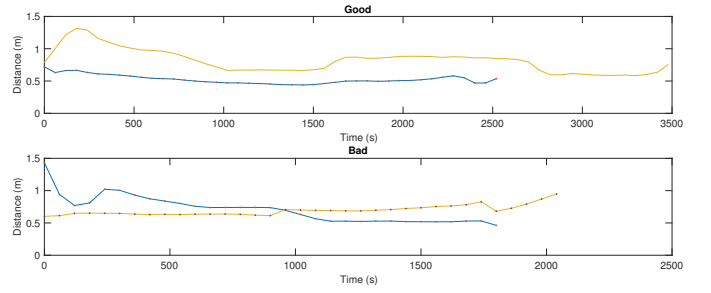ull set of features with a given selection of the other parameters mentioned above and we proceed removing features, as a form of model selection.

Regarding the effects of the multimodal learning analytics features on predicting students' group performances, below results are found:

- IDEC (Arduino IDE) removal does not effect the results of the classifiers,

- Removal of all face and hand duration has very little effect on the classifiers,

- Distance measures DHB and DBL **alone** are capable of predicting the results with a high accuracy (0.75) across classifiers,

- The audio level feature AUD **alone** is currently a **strong** feature for classification (1.0 with Naive Bayes) with time windows 5min,10min and 30min and binary scoring.

Interestingly the logistic regression is capable of an optimal result (1.0) when considering IDEX, IDEVHW, IDEVSW, and DBL, which are the main IDE features, except component counts and the distance between learners (DBL).

VI. DISCUSSION

In this paper, we processed and extracted multimodal learning analytics features to identify which features of students' group work are good predictors of successful project outcomes in open-ended tasks with physical computing. As an independent variable we used human judgment of the quality of students' artefact. We initially focused on identifying the different phases of work in relation to accuracy in predicting the group success. We found that the planning and building stages of students learning activities are better predictors of their success than the reflection stage. Then, we investigated the certain features of the MMLA, in order to determine which features can predict the group success with higher accuracy. Our results show that the Distance between Hands **DBH** and Distance between Learners **DBL** are key features to predict students' performances in practice-based learning activities. In our case, they highly correlate with the quality of the students' artefacts in practice-based learning. These results are aligned with existing MMLA research findings that show the potential of hand motion and speed, and the location of the learners to predict student success at various learning outcomes. [9], [8], [11]. Figure 3 shows the distance along time in four sessions.

The other features of MMLA such as Hand Motion Speed **HMS**, Faces Looking at the Screen **FLS**, did not perform very well to predict students' artefact quality across the classifiers. While the Arduino IDE the Number of Active Blocks **IDEC**, the Variety of Hardware **IDEVHW** and Software Blocks used **IDEVSW** and the number of interconnections between blocks as a Measure of Complexity **IDEX** were able to predict students' outcomes, they were only marginal across the classifiers. Furthermore, the audio signal level **AUD** appears to be a promising feature to predict performance, however more investigation is needed for using this feature in combination with others.

One of the main limitations of our approach is on the scoring of the sessions that is limited to a binary classification with respect to a richer 3-level human scoring. In our future research, we plan to extend observer scale so that it comprises more levels of judgment or multiple criteria for student inter-action actions to create the training sets.

VII. CONCLUSION

In this paper, we started from the hypothesis that specific features in multimodal learning analytics can provide useful information about the quality of groups' artefacts in practice-based learning. From the high-frequency multimodal data collected, we compared different machine learning classifiers in their accuracy to predict human grading of the groups' artefacts. Then, using these classifiers, we identified the most effective features of MMLA to predict the students' group performances in practice-based learning activities. In the next stage of work, we plan to use these optimal features to train a General Linear Model Regression model that can provide statistical evidence about the estimation process for the grading of students' group design outcomes automatically.

REFERENCES

[1] D. Clow, "An overview of learning analytics," *Teaching in Higher Education*, vol. 18, no. 6, pp. 683–695, aug 2013. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/13562517.2013.827653

[2] G. Siemens and R. S. J. d. Baker, "Learning analytics and educational data mining: Towards communication and collaboration," in *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12*. New York, New York, USA: ACM Press, apr 2012, CONFERENCE_PAPER, p. 252. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2330601.2330661

[3] E. R. Halverson and K. Sheridan, "The maker movement in education," *Harv Educ Rev*, vol. 84, no. 4, pp. 495–504, dec 2014. [Online]. Available: http://hepgjournals.org/doi/10.17763/haer.84.4.34j1g68140382063

[4] P. Blikstein and M. Worsley, "Multimodal learning analytics and education data mining: using computational technologies to measure complex learning tasks," *Journal of Learning Analytics*, sep 2016. [Online]. Available: http://epress.lib.uts.edu.au/journals/index.php/JLA/article/view/4383/5596

[5] P. Mulholland, S. Anastopoulou, T. Collins, M. Feisst, M. Gaved, L. Kerawalla, M. Paxton, E. Scanlon, M. Sharples, and M. Wright, "nquire: Technological support for personal inquiry learning," *IEEE Trans. Learning Technol.*, vol. 5, no. 2, pp. 157–169, apr 2012. [Online]. Available: http://ieeexplore.ieee.org/document/6095502/

[6] X. Ochoa and M. Worsley, "Editorial: Augmenting learning analytics with multimodal sensory data," *Journal of Learning Analytics*, sep 2016. [Online]. Available: http://epress.lib.uts.edu.au/journals/index.php/JLA/article/view/5081/5595

[7] P. Blikstein, "Multimodal learning analytics," in *Proceedings of the Third International Conference on Learning Analytics and Knowledge - LAK '13*, D. Suthers and K. Verbert, Eds. New York, New York, USA: ACM Press, apr 2013, CONFERENCE_PAPER, p. 102. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2460296.2460316

[8] X. Ochoa, K. Chiluiza, G. Mndez, G. Luzardo, B. Guamn, and J. Castells, "Expertise estimation based on simple multimodal features," in *Proceedings of the 15th ACM on International conference on multimodal interaction - ICMI '13*. New York, New York, USA: ACM Press, dec 2013, CONFERENCE_PAPER, pp. 583–590. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2522848.2533789

[9] P. Blikstein, "Using learning analytics to assess students' behavior in open-ended programming tasks," in *Proceedings of the 1st International Conference on Learning Analytics and Knowledge - LAK '11*. New York, New York, USA: ACM Press, feb 2011, CONFERENCE_PAPER, p. 110. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2090116.2090132

[10] M. Worsley and P. Blikstein, "Analyzing engineering design through the lens of computation," *Journal of Learning Analytics*, vol. 1, no. 2, pp. 151–186, aug 2014.

[11] S. Grover, M. Bienkowski, A. Tamrakar, B. Siddiquie, D. Salter, and A. Divakaran, "Multimodal analytics to study collaborative problem solving in pair programming," in *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*. New York, New York, USA: ACM Press, apr 2016, CONFERENCE_PAPER, pp. 516–517. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2883851.2883877

[12] L. P. Prieto, K. Sharma, P. Dillenbourg, and M. Jess, "Teaching analytics: Towards automatic extraction of orchestration graphs using wearable sensors," in *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*. New York, New York, USA: ACM Press, apr 2016, CONFERENCE_PAPER, pp. 148–157. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2883851.2883927

[13] R. Martinez-Maldonado, B. Schneider, S. Charleer, S. B. Shum, J. Klerkx, and E. Duval, "Interactive surfaces and learning analytics: Data, orchestration aspects, pedagogical uses and challenges," in *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*. New York, New York, USA: ACM Press, apr 2016, CONFERENCE_PAPER, pp. 124–133. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2883851.2883873

[14] E. Ruffaldi, G. Dabisias, L. Landolfi, and D. Spikol, "Data collection and processing for a multimodal learning analytic system," in *2016 SAI Computing Conference (SAI)*. IEEE, jul 2016, CONFERENCE_PAPER, pp. 858–863. [Online]. Available: http://ieeexplore.ieee.org/document/7556079/

[15] C. J. Atman, R. S. Adams, M. E. Cardella, J. Turns, S. Mosborg, and J. Saleem, "Engineering design processes: A comparison of students and expert practitioners," *J. Eng. Educ.*, vol. 96, no. 4, pp. 359–379, oct 2007. [Online]. Available: http://doi.wiley.com/10.1002/j.2168-9830.2007.tb00945.x

[16] M. Cukurova, K. Avramides, D. Spikol, R. Luckin, and M. Mavrikis, "An analysis framework for collaborative problem solving in practice-based learning activities: A mixed-method approach," in *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*. New York, New York, USA: ACM Press, apr 2016, CONFERENCE_PAPER, pp. 84–88. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2883851.2883900

[17] S. Kotsiantis, K. Patriarcheas, and M. Xenos, "A combinational incremental ensemble of classifiers as a technique for predicting students performance in distance education," *Knowledge-Based Systems*, vol. 23, no. 6, pp. 529–535, 2010.

[18] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.