




[Center for Machine Learning and Intelligent Systems](#)

[About](#) [Citation Policy](#) [Donate a Data Set](#) [Contact](#)

☒ Repository ☐ Web 

[View ALL Data Sets](#)

Educational Process Mining (EPM): A Learning Analytics Data Set Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Educational Process Mining data set is built from the recordings of 115 subjects' activities through a logging application while learning with an educational simulator.

Data Set Characteristics:	Multivariate, Sequential, Time-Series	Number of Instances:	230318	Area:	Computer
Attribute Characteristics:	Integer	Number of Attributes:	13	Date Donated	2015-09-24
Associated Tasks:	Classification, Regression, Clustering	Missing Values?	N/A	Number of Web Hits:	40583

Source:

Mehmoosh Vahdat(1,2), Luca Oneto(1), Davide Anguita(1), Mathias Funk (2), and Matthias Rauterberg (2)

1 - Smartlab - Non-Linear Complex Systems Laboratory

DIBRIS - Università degli Studi di Genova, Genoa (I-16145), Italy.

2 - Department of Industrial Design, Eindhoven University of Technology, 5612AZ Eindhoven, The Netherlands

la '@' smartlab.ws

Data Set Information:

The experiments have been carried out with a group of 115 students of first-year, undergraduate Engineering major of the University of Genoa.

We carried out this study over a simulation environment named Deeds (Digital Electronics Education and Design Suite) which is used for e-learning in digital electronics. The environment provides learning materials through specialized browsers for the students, and asks them to solve various problems with different levels of difficulty. For more information about the Deeds simulator used for this course look at: [\[Web Link\]](#)

and to know more about the exercises contents of each session see 'exercises_info.txt'.

Our data set contains the students' time series of activities during six sessions of laboratory sessions of the course of digital electronics. There are 6 folders containing the students' data per session. Each 'Session' folder contains up to 99 CSV files each dedicated to a specific student log during that session. The number of files in each folder changes due to the number of students present in each session. Each file contains 13 features. See 'features_info.txt' for more details.

For the details of activities performed by the students during the course, see 'activities_info.txt'

The data set includes the following files:

=====

- 'README.txt'
- 'features_info.txt': contains information about the variables used on the feature vector.
- 'features.txt': List of all features.
- 'activities_info.txt': contains information about the variable 'activity'.
- 'activities.txt': list of all activities.
- 'exercises_info.txt': contains information about the variable 'exercise'.
- 'grades_info.txt': contains information about the grade data.

Data:

=====

- 'Processes': contains the data files from Session 1 to 6.
- 'logs.txt': shows information about the log data per student Id. It shows whether a student has a log in each session (0: has no log, 1: has log).
- 'final_grades.xlsx': contains the results of the final exam in two sheets.
- 'intermediate_grades.xlsx': contains the grades for the students' assignments per session.
- 'final_exam.pdf': shows the content of the final exam (original in Italian).
- 'final_exam_ENG.pdf': shows the content of the final exam translated in English.

Notes:

=====

For more information about this data set please look at:

www.la.smartlab.ws

la '@' smartlab.ws

Attribute Information:

The features selected for this data set come from pre-processing of data collected through a logging program.

Due to ethical reasons and to ensure the anonymity of our users, we cannot share the original log files, instead, we share the data transformed and cleaned in an appropriate format.

The original logs contain the logging data of client system per approximately a second, while the features are calculated in order to be allocated to a particular activity.

The features are selected and presented in a suitable format for Process Mining. In this sense, the data is presented per session, per student, and per exercise. Each CSV file belongs to a specific session and a specific student (named by the student Id). Each file contains several exercises of that session presented in 'exercise' feature. Each 'exercise' contains activities, which start-time, end-time, and other features are allocated to that.

For further information about each feature, see 'features_info.txt'.

Relevant Papers:

M. Vahdat, L. Oneto, A. Ghio, G. Donzellini, D. Anguita, M. Funk, M. Rauterberg.: A learning analytics methodology to profile students behavior and explore interactions with a digital electronics simulator. In: de Freitas, S., Rensing, C., Ley, T., Munoz-Merino, P.J. (eds.) EC-TEL 2014. LNCS, vol. 8719, pp. 596â€“597. Springer (2014).

M. Vahdat, A. Ghio, L. Oneto, D. Anguita, M. Funk, M. Rauterberg, Advances in learning analytics and educational data mining, in: European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (2015).

Citation Request:

M. Vahdat, L. Oneto, D. Anguita, M. Funk, M. Rauterberg.: A learning analytics approach to correlate the academic achievements of students with interaction data from an educational simulator. In: G. Conole et al. (eds.): EC-TEL 2015, LNCS 9307, pp. 352-366. Springer (2015).
DOI: 10.1007/978-3-319-24258-3_26

Supported By:



In Collaboration With:



[About](#) || [Citation Policy](#) || [Donation Policy](#) || [Contact](#) || [CML](#)