

Students' Academic Performance Dataset

Data Set Characteristics: Multivariate

Number of Instances: 480

Area: E-learning, Education, Predictive models, Educational Data Mining

Attribute Characteristics: Integer/Categorical

Number of Attributes: 16

Date: 2016-11-8

Associated Tasks: Classification

Missing Values? No

File formats: xAPI-Edu-Data.csv

Source:

Elaf Abu Amrieh, Thair Hamtini, and Ibrahim Aljarah, The University of Jordan, Amman, Jordan, <http://www.Ibrahimaljarah.com> www.ju.edu.jo

Dataset Information:

This is an educational data set which is collected from learning management system (LMS) called Kalboard 360. Kalboard 360 is a multi-agent LMS, which has been designed to facilitate learning through the use of leading-edge technology. Such system provides users with a synchronous access to educational resources from any device with Internet connection.

The data is collected using a learner activity tracker tool, which called experience API (xAPI). The xAPI is a component of the training and learning architecture (TLA) that enables to monitor learning progress and learner's actions like reading an article or watching a training video. The experience API helps the learning activity providers to determine the learner, activity and objects that describe a learning experience. The dataset consists of 480 student records and 16 features. The features are classified into three major categories: (1) Demographic features such as gender and nationality. (2) Academic background features such as educational stage, grade Level and section. (3) Behavioral features such as raised hand on class, opening resources, answering survey by parents, and school satisfaction.

The dataset consists of 305 males and 175 females. The students come from different origins such as 179 students are from Kuwait, 172 students are from Jordan, 28 students from Palestine, 22 students are from Iraq, 17 students from Lebanon, 12 students from Tunis, 11 students from Saudi Arabia, 9 students from Egypt, 7 students from Syria, 6 students from USA, Iran and Libya, 4 students from Morocco and one student from Venezuela.

The dataset is collected through two educational semesters: 245 student records are collected during the first semester and 235 student records are collected during the second semester.

The data set includes also the school attendance feature such as the students are classified into two categories based on their absence days: 191 students exceed 7 absence days and 289 students their absence days under 7.

This dataset includes also a new category of features; this feature is parent participation in the educational process. Parent participation feature have two sub features: Parent Answering Survey and Parent School Satisfaction. There are 270 of the parents answered survey and 210 are not, 292 of the parents are satisfied from the school and 188 are not.

(See the related papers for more details).

Attributes

1 Gender - student's gender (nominal: 'Male' or 'Female')

2 Nationality- student's nationality (nominal: 'Kuwait', 'Lebanon', 'Egypt', 'SaudiArabia', 'USA', 'Jordan', 'Venezuela', 'Iran', 'Tunis', 'Morocco', 'Syria', 'Palestine', 'Iraq', 'Lybia')

3 Place of birth- student's Place of birth (nominal: 'Kuwait', 'Lebanon', 'Egypt', 'SaudiArabia', 'USA', 'Jordan', 'Venezuela', 'Iran', 'Tunis', 'Morocco', 'Syria', 'Palestine', 'Iraq', 'Lybia')

4 Educational Stages- educational level student belongs (nominal: 'lowerlevel', 'MiddleSchool', 'HighSchool')

5 Grade Levels- grade student belongs (nominal: 'G-01', 'G-02', 'G-03', 'G-04', 'G-05', 'G-06', 'G-07', 'G-08', 'G-09', 'G-10', 'G-11', 'G-12')

6 Section ID- classroom student belongs (nominal: 'A', 'B', 'C')

7 Topic- course topic (nominal: 'English', 'Spanish', 'French', 'Arabic', 'IT', 'Math', 'Chemistry', 'Biology', 'Science', 'History', 'Quran', 'Geology')

8 Semester- school year semester (nominal: 'First', 'Second')

9 Parent responsible for student (nominal: 'mom', 'father')

10 Raised hand- how many times the student raises his/her hand on classroom (numeric: 0-100)

11- Visited resources- how many times the student visits a course content (numeric: 0-100)

12 Viewing announcements-how many times the student checks the new announcements (numeric: 0-100)

13 Discussion groups- how many times the student participate on discussion groups (numeric: 0-100)

14 Parent Answering Survey- parent answered the surveys which are provided from school or not (nominal: 'Yes', 'No')

15 Parent School Satisfaction- the Degree of parent satisfaction from school (nominal: 'Yes', 'No')

16 Student Absence Days- the number of absence days for each student (nominal: above-7, under-7)

The students are classified into three numerical intervals based on their total grade/mark:

Low-Level: interval includes values from 0 to 69,

Middle-Level: interval includes values from 70 to 89,

High-Level: interval includes values from 90-100.

Relevant Papers:

- Amrieh, E. A., Hamtini, T., & Aljarah, I. (2016). Mining Educational Data to Predict Student's academic Performance using Ensemble Methods. *International Journal of Database Theory and Application*, 9(8), 119-136.
- Amrieh, E. A., Hamtini, T., & Aljarah, I. (2015, November). Preprocessing and analyzing educational data set using X-API for improving student's performance. In *Applied Electrical Engineering and Computing Technologies (AEECT)*, 2015 IEEE Jordan Conference on (pp. 1-5). IEEE.

Citation Request:

Please include these citations if you plan to use this dataset:

- Amrieh, E. A., Hamtini, T., & Aljarah, I. (2016). Mining Educational Data to Predict Student's academic Performance using Ensemble Methods. *International Journal of Database Theory and Application*, 9(8), 119-136.
- Amrieh, E. A., Hamtini, T., & Aljarah, I. (2015, November). Preprocessing and analyzing educational data set using X-API for improving student's performance. In *Applied Electrical Engineering and Computing Technologies (AEECT)*, 2015 IEEE Jordan Conference on (pp. 1-5). IEEE.