

Feature name	Type
total_time_spent	Numerical
average_gap	Numerical
longest_gap	Numerical
total_modules	Numerical
total_child_modules	Numerical
unique_module_categories	Numerical
average_children_per_module	Numerical
has_problem	Categorical (Binary)
has_video	Categorical (Binary)
has_discussion	Categorical (Binary)
most_active_period_Morning	Categorical (Binary)
most_active_period_Afternoon	Categorical (Binary)
most_active_period_Evening	Categorical (Binary)
most_active_period_Night	Categorical (Binary)
activity_decline	Numerical

Table 1. Sample schema of learner features.

File name	Description
enrollment_train.csv	Training set enrolment records with <i>enrollment_id</i> , <i>username</i> , and <i>course_id</i> used to link logs and ground truth.
enrollment_test.csv	Test set enrolment records with the same keys as training, without labels.
log_train.csv	Timestamped event logs for each training enrolment, including <i>time</i> , <i>source</i> , <i>event</i> , and <i>object</i> .
log_test.csv	Timestamped event logs for test enrolments, same schema as training logs.
object.csv	Course object catalogue that maps <i>object</i> identifiers to hierarchical course structure and resource types.
truth_train.csv	Ground truth labels for training enrolments, indicating whether the learner dropped out by the end of the course.

Table 2. KDD Cup 2015 files and brief descriptions.

File name	Description
studentInfo.csv	Demographic and prior education attributes for each student and presentation, including age band, gender, highest education, region, disability, and deprivation index.
studentRegistration.csv	Registration and withdrawal dates for each student and module presentation, used to infer enrolment status and study outcome.
courses.csv	Module and presentation metadata, such as module code, presentation code, start date, and presentation length.
assessments.csv	Definitions of assessments for each module presentation, including type, weight, and scheduling information.
studentAssessment.csv	Student level assessment records, including submitted attempts, marks, lateness, and missing submissions.
vle.csv	Catalogue of virtual learning environment resources, with resource identifiers, activity types, and URLs.
studentVle.csv	Daily interaction logs per student and resource, providing click counts by date across a module presentation.

Table 3. OULAD files and brief descriptions.

Dataset	Class	Before	After
Train	Continue (0)	24,961	95,581
	Dropout (1)	95,581	95,581
Test	Continue (0)	16,720	63,642
	Dropout (1)	63,642	63,642

Table 4. Class distribution before and after GAN-based balancing.

True label	Predicted Stayed	Predicted Dropped
Stayed	55,003	8,639
Dropped	2,512	61,130

Table 5. Confusion Matrix– VertiDaX Model