

The description of the training and test data sets is as follows:

#### Training Data:

Number of instances: 72326

1st column is enrollment ID, which corresponds to a user's enrollment in a course

2nd column is label

The rest of the columns are features

#### Test Data:

Number of instances: 48216

1st column is enrollment ID (Col 0)

The rest of the columns are features

#### Features:

Number of features: 50

##### Course related features:

1. Number of modules in the course
2. Number of chapters in the course
3. Number of problems in the course
4. Number of videos in the course
5. Number of discussions in the course
6. Number of peergrading in the course
7. Number of users who enrolled the course
8. Number of activities from users enrolled the course
9. The ratio between number of chapters accessed by users and the total number of chapters in the course
10. Number of Dropout for the course
11. Max duration among all users
12. Average duration for the course
13. Min number of days spent on the course among all users
14. Average number of days spent on the course

##### User related features:

15. Number of Courses enrolled by the user
16. Min number of activities per day
17. Max number of activities per day
18. Average number of activities per day
19. Max duration among all courses enrolled by the user
20. Min duration among all courses enrolled by the user
21. Average duration over all courses enrolled by the user

22. Max number of days spent on a course
23. Min number of days spent on a course
24. Average number of days spent on a course

Enrollment related features:

25. Max interval between days when there are activities
26. Min interval between days when there are activities
27. Average interval between days when there are activities
28. Min number of activities per day
29. Max number of activities per day
30. Average number of activities per day
31. Duration of the enrollment
32. Number of activities
33. Number of spent days
34. The ratio between the number of accesses in the Morning and the total number of accesses.
35. The ratio between the number of accesses in the Evening and the total number of accesses.
36. The ratio between the number of accesses at Night and the total number of accesses.
37. The ratio between number of chapters accessed by the user and the total number of chapters in the course
38. The ratio between number of problems accessed by the user and the total number of problems in the course
39. The ratio between number of videos accessed by the user and the total number of videos in the course
40. Times of visiting wiki
41. Times of participating discussion
42. Number of activities during week days
43. Number of activities during week ends
44. Time difference in days between the start data of the course and the first day the user accessed the course
45. Max number of consecutive days
46. The ratio between the number of activities for the enrollment and the average number of activities for the course
47. The average length of time spent on the course per day
48. The max length of time spent on the course per day
49. The min length of time spent on the course per day
50. Time difference in days between the end data of the course and the last day the user accessed the course