My thoughts on the data

I agree with Siim that there is a lot of information in these files, and we need to figure out the most simple and logical way to analyze the log records and entries from the gradebook.

In my opinion, from log records we could use Aeg, Kasutaja täisnimi, Sündmuse kontekst, Sündmuse nimi. It seems to me that the only thing we can calculate using the feature Aeg is the time spent to solve the tests and maybe count of nighttime views. As Sündmuse nimi doesn't contain for example test numbers it might be reasonable aggregate a new field from Sündmuse kontekst and Sündmuse nimi.

To profile typical students, we could monitor at what time of the day they take tests and submit assignments; the average time to perform a test; and we could also use the number of total logs, count of nighttime views; average test and assignment result, results of kontrolltöö1, kontrolltöö2 and the exam.

Identifying struggling students we could classification model based on:

1. How often a student visits Moodle (measured by the number of total logs)

2. What are the results of the student's tests

3. What are the results of the student's homework

I am pretty confused about predicting the final grade based on the first weeks activity as 70% of the final result comes from kontrolltöö1, kontrolltöö2 and the final exam. Student can be very motivated and active in first weeks but if he/she fails on kontrolltöö1, kontrolltöö2 and the final exam he/she can't get a positive result.