

Educational Process Mining:
Monitoring student academic
outcome from process perspective

Research map

Overall RQ: How can monitoring the learning process of students from Learning Management Systems support to improve their academic performance?

Sub objectives	Predict student's outcome based on their learning event logs	Understanding learning process: Process discovery	Learning process comparison between successful and less-successful student cohorts	Provide students with feedback on learning process
Methodology	Machine Learning approaches on event logs <ul style="list-style-type: none"> - PCA (to reduce dimensionality where necessary) - Random Forest - Regressions - Multi Layers Neural Nets - XGBoost 	Process Discovery approaches: <ul style="list-style-type: none"> - Heuristics Miner - Inductive Miner - DFG Miner - Alpha Miner - FCA Miner (proposed to deal with the complexity of the process) 	Conformance checking approaches <ul style="list-style-type: none"> - Replay token - Footprint - Process model (Petri Nets) comparison: <ul style="list-style-type: none"> + Control flow + Materials used throughout the process 	Provide personal feedback profile for students regarding learning process: <ul style="list-style-type: none"> + Suggest successful learning process + Emphasize what to do more Test on real courses Compare student's results Get feedback from students & educators
Validation & Evaluation	Prediction scores: <ul style="list-style-type: none"> - Cross validation - Accuracy - Precision - Recall - F1 score 	Evaluation resulting process models <ul style="list-style-type: none"> - Replay fitness - Precision - Generalisation - Simplicity 		<ul style="list-style-type: none"> - Test on real programming classes (expected - semester 1 - 2020/2021) - Compare student's results - Get feedback from students & educators about the notification system.
Impact	Support learners through Learning Management Systems	Process mining applications in combination with machine learning in educational domain	Help educators understand learner's behaviours	

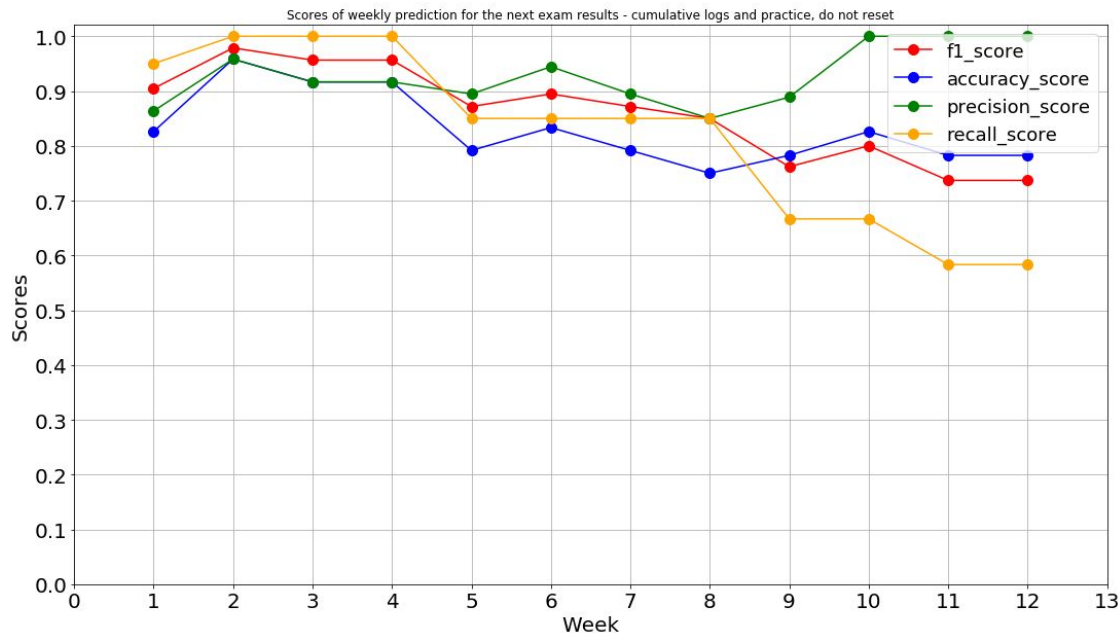
Learning prediction based on learning event log

Features (trained and validated per week):

- The frequency of learning activities on course materials per week observed from Einstein: (load, scroll, blur, remain, left-click, right-click, select keyword, change section, unload)
- Cumulative practice result (correct submission/total submission)
- Target variable: Examination result at week 4,8,12 - Pass (≥ 0.4) and Failed
 - Week 1-2-3-4 data predicts exam results on week 4;
 - Week 5-6-7-8 - Exam on Week 8
 - Week 9-10-11-12 - Exam on W12

Techniques: Random forest (so far)

Findings: Those features can be used to predict student outcome in programming classes. The prediction performances are very good across 12 weeks.



Learning process discovery - Experimental results

	Proposed method (FCA Miner)	DFG Miner - Tool	Inductive Miner - Tool
Log 1.1 Exam1	Fitness: 0.62 Precision: 0.93 Simplicity: 0.61 Generalisation: 0.54	Fitness: 0.48 Precision: 1 Simplicity: 0.87 Generalisation: 0.13	Fitness: 0.54 Precision: 0.60 Simplicity: 0.52 Generalisation: 0.97
	0.68/0.74	0.62/0.65	0.66/0.57
Log 1.2 Exam1	Fitness: 0.54 Precision: 0.89 Simplicity: 0.63 Generalisation: 0.5	Fitness: 0.5 Precision: 1 Simplicity: 0.87 Generalisation: 0.13	Fitness: 0.69 Precision: 0.47 Simplicity: 0.54 Generalisation: 0.96
	0.64/0.67	0.63/0.66	0.67/0.56
Log 2.1 Exam2	Fitness: 0.41 Precision: 0.92 Simplicity: 0.74 Generalisation: 0.32	Fitness: 0.54 Precision: 1 Simplicity: 0.87 Generalisation: 0.13	Fitness: 0.55 Precision: 0.56 Simplicity: 0.52 Generalisation: 0.93
	0.6/0.58	0.63/0.7	0.64/0.55
Log 2.2 Exam2	Fitness: 0.73 Precision: 0.93 Simplicity: 0.67 Generalisation: 0.43	Fitness: 0.45 Precision: 0.99 Simplicity: 0.87 Generalisation: 0.13	Fitness: 0.57 Precision: 0.51 Simplicity: 0.52 Generalisation: 0.97
	0.69/0.82	0.61/0.62	0.64/0.54
Log 3.1 Exam3	Fitness: 0.47 Precision: 0.92 Simplicity: 0.74 Generalisation: 0.30	Fitness: 0.35 Precision: 1 Simplicity: 0.87 Generalisation: 0.13	Fitness: 0.5 Precision: 0.53 Simplicity: 0.5 Generalisation: 0.95
	0.64	0.59	0.62
Log 3.2 Exam3	Fitness: 0.54 Precision: 0.96 Simplicity: 0.68 Generalisation: 0.4	Fitness: 0.43 Precision: 1 Simplicity: 0.87 Generalisation: 0.14	Fitness: 0.67 Precision: 0.56 Simplicity: 0.53 Generalisation: 0.92
	0.64	0.60	0.67

* Alpha Miner and Heuristics Miner cannot produce a Workflow-nets with these event logs (Petri Nets with one input and one output, and other places and transitions are on the path connecting input and output)

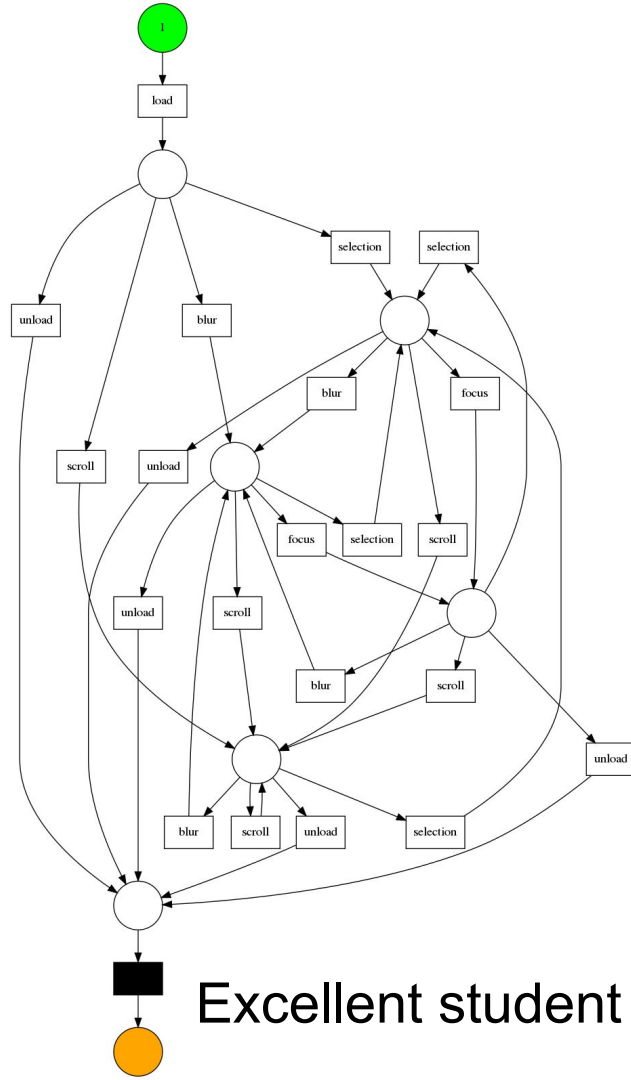
Learning process discovery - Experimental results

	Proposed method	DFG Miner	Inductive Miner
Log 1.1 Exam1	Number of places: 7 Number of transitions: 11 Number of arcs: 22	Number of places: 12 Number of transitions: 82 Number of arcs: 164	Number of places: 41 Number of transitions: 43 Number of arcs: 98
Log 1.2 Exam1	Number of places: 7 Number of transitions: 12 Number of arcs: 24	Number of places: 13 Number of transitions: 83 Number of arcs: 166	Number of places: 30 Number of transitions: 35 Number of arcs: 78
Log 2.1 Exam2	Number of places: 9 Number of transitions: 25 Number of arcs: 50	Number of places: 13 Number of transitions: 85 Number of arcs: 170	Number of places: 42 Number of transitions: 45 Number of arcs: 102
Log 2.2 Exam2	Number of places: 7 Number of transitions: 14 Number of arcs: 28	Number of places: 12 Number of transitions: 81 Number of arcs: 162	Number of places: 42 Number of transitions: 45 Number of arcs: 102
Log 3.1 Exam3	Number of places: 9 Number of transitions: 26 Number of arcs: 52	Number of places: 12 Number of transitions: 83 Number of arcs: 166	Number of places: 44 Number of transitions: 45 Number of arcs: 104
Log 3.2 Exam3	Number of places: 7 Number of transitions: 15 Number of arcs: 30	Number of places: 12 Number of transitions: 79 Number of arcs: 158	Number of places: 44 Number of transitions: 50 Number of arcs: 110

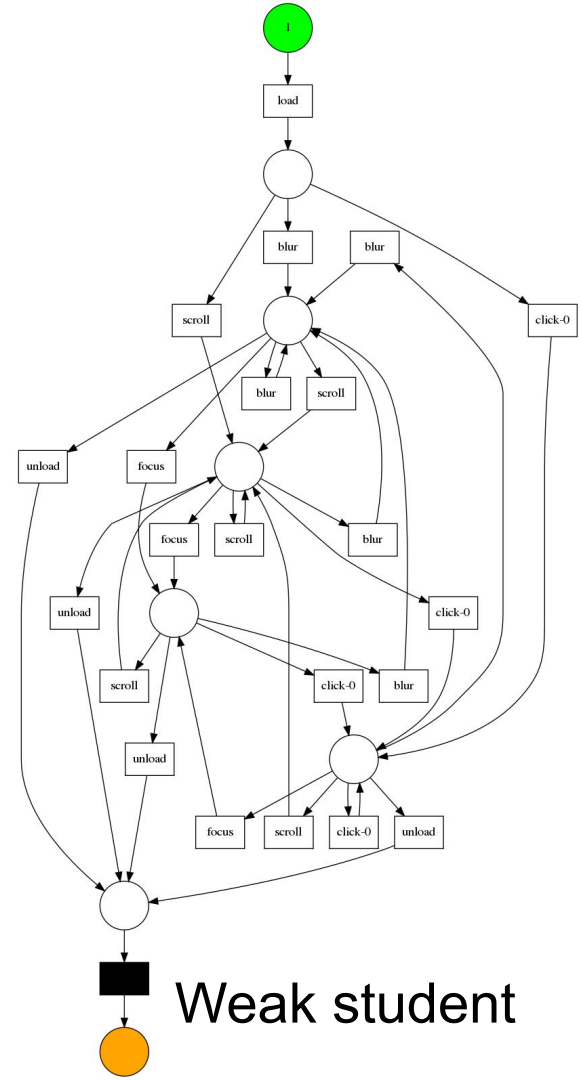
The proposed method produces process models (workflow-nets) with the same quality in comparison with other methods.

The resulting models from the method contain a fewer number of elements than the models from other methods to describe the same event log.

Proposed method (FCA Miner)

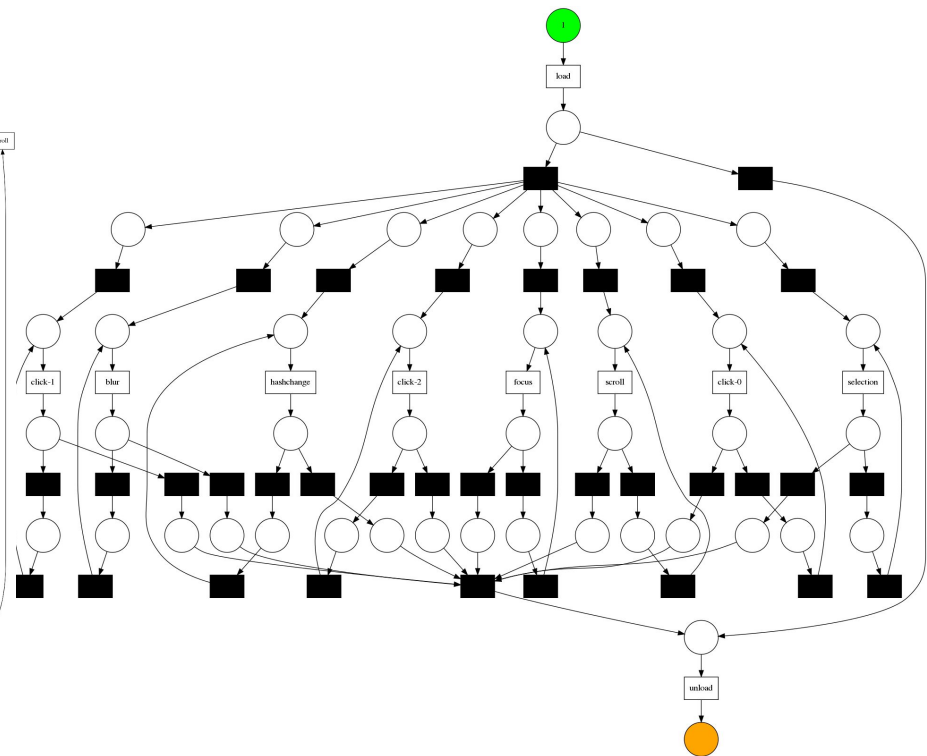
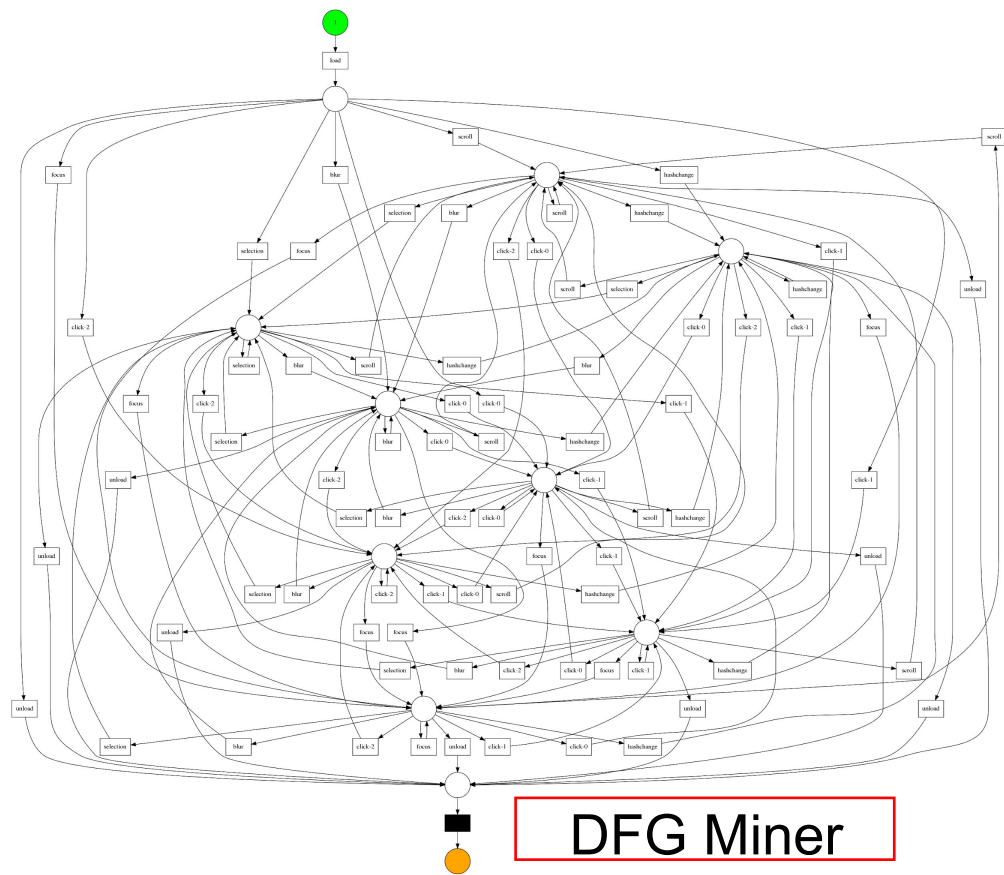


Excellent student



Weak student

“Other methods”



Inductive Miner

Personal Feedback Profile for students

1. The learning process model from a successful cohort in the class
2. The student's learning process model
3. The comparison between the two models regarding:
 - a. Learning control flow
 - b. Materials used throughout the process
4. Suggestion for students:
 - a. Learning control flow: Keep in mind to do what and what when studying on LMS
 - b. Materials used in each steps

Student learning behaviour profile for educators

A dashboard to navigate learning processes from students in the course:

1. Learning process models from a student/cohort.
2. Compare learning process models.

Problem

Prediction is very good at the beginning, but worse in the last 4 weeks of the semester (for exam 3) - but the scores are still better than previous work :-?

Resulting learning process model are hard to understand for users (educators & students)