Prédire l‘abandon des études : Comment s‘y prendre ?

BHT
Berliner Hochschule für Technik

16-11-2022
Agathe Merceron

Studiere Zukunft

Berlin

- A nice place to live!

http://www.iheartberlin.de/berlin-is-on-ice-impressions-of-a-frozen-city/
http://awesomeberlin.net/wp-content/uploads/2017/06/wann1.jpg

Démarche centrée données et utilisatrices / utilisateurs

- Acquisition, stockage et utilisation de données éducationnelles administratives
- Algorithmes de prédiction
- Problèmes d’équité
- Utilisation possible des résultats

Project Students Advice

- https://projekt.bht-berlin.de/students-advice/

Petra Sauer

Kerstin Wagner

Project Students Advice

- Work began in 2019 – In this talk: mainly stand in 2022.

Lennart Egbers

Stephan Wagner

Daria Novoseltseva

Predicting Dropout of a Degree: which Data?

- German higher education
- Three degree programs: six-semesters bachelor
- Data from 1809 students, 1007 with the label graduate and 802 with the label dropout
- Data from 2012 till 2019:
  - Enrollment date in the degree program
  - Courses enrolled with marks and semester
  - Graduation date or exmatriculation date
  - Gender

Predicting Dropout of a Degree: how to obtain the data?

- Explain the project to the Data Privacy Officer
- Understand what is possible taking into account GDPR
- Write a Data Security Concept
- Explain the project to the vice-president for Teaching and Learning
- Put in place a procedure to obtain the data from the administration

Predicting Dropout of a Degree: Example Data

Prog	ID	Gender	Status	Start	Exam_Sem	Sem	Module_ID	Module	Plan_Sem	Grade	Label
B-DMT	362	W	Graduate	4029	4031	3	WP26	Fotografie	34	nan	Enrolled
B-MI	5618	W	Dropout	4036	4036	1	B04	Technische Grundlagen der Informatik	1	nan	Enrolled
B-ARCH	3964	M	Active	4037	4037	1	B05	Baugeschichte und Architekturlehre	1	1.7	Passed
B-ARCH	7218	W	Graduate	4031	4034	4	B21	Entwerfen und Konstruieren im Bestand	4	2.3	Passed
B-ARCH	460	W	Active	4034	4037	4	B15	Entwerfen und Konstruieren	3	5.0	Failed

Students Grades Example (StAd)

Predicting Dropout of a Degree: What can be wrong in the Data?

- Date of immatriculation?
- Marks?
- ….

➢ Data Cleaning

Predicting Dropout of a Degree: Features and Algorithms

- Features:
  - Socio-economic.
  - Performance:
    - Local features: marks in courses, etc. specific to a study program
    - Global features: average mark, number of courses passed, etc. independent of a specific study program

➢ R. Manrique, B. P. Nunes, O. Marino, M. A. Casanova, and T. Nurmikko-Fuller, 2019 An analysis of student representation, representative features and classification algorithms to predict degree dropout. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, pages 401-410. ACM, 2019.

Predicting Dropout of a Degree: Algorithms

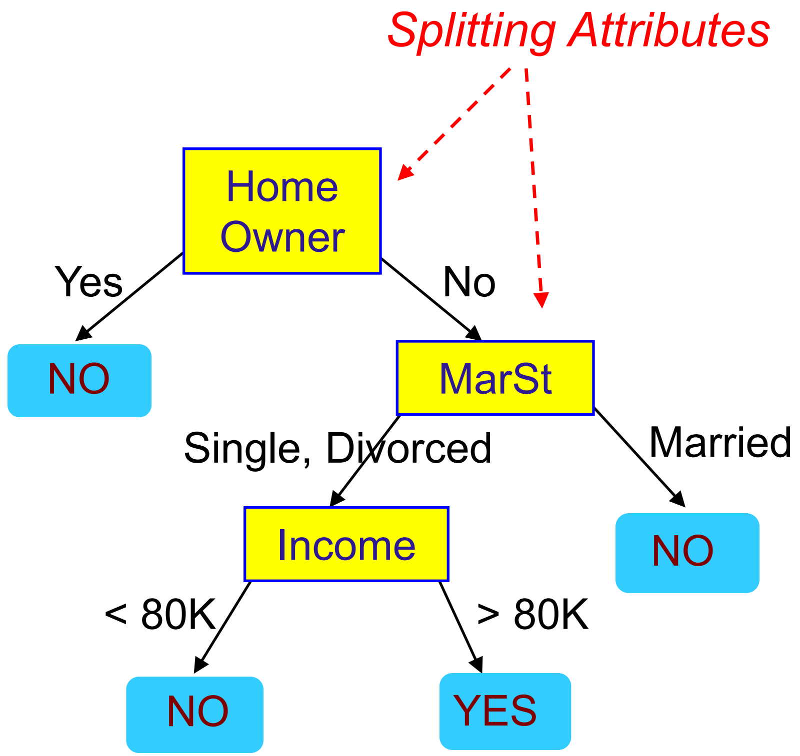
- Removal of different kinds of outliers.
- Different algorithms:
  - Decision trees (DT) – explainable.
  - Logistics regression (LR) – explainable.
  - Random Forest (RF) – Ensemble method; might give better results.
  - Neural Networks (NN) – usually non explainable.
  - …..
- Data split into training set and test set:
  - Models built on the training set and evaluated on the test set (time aware).

Example of a Decision Tree

categorical categorical continuous class

ID	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Training Data

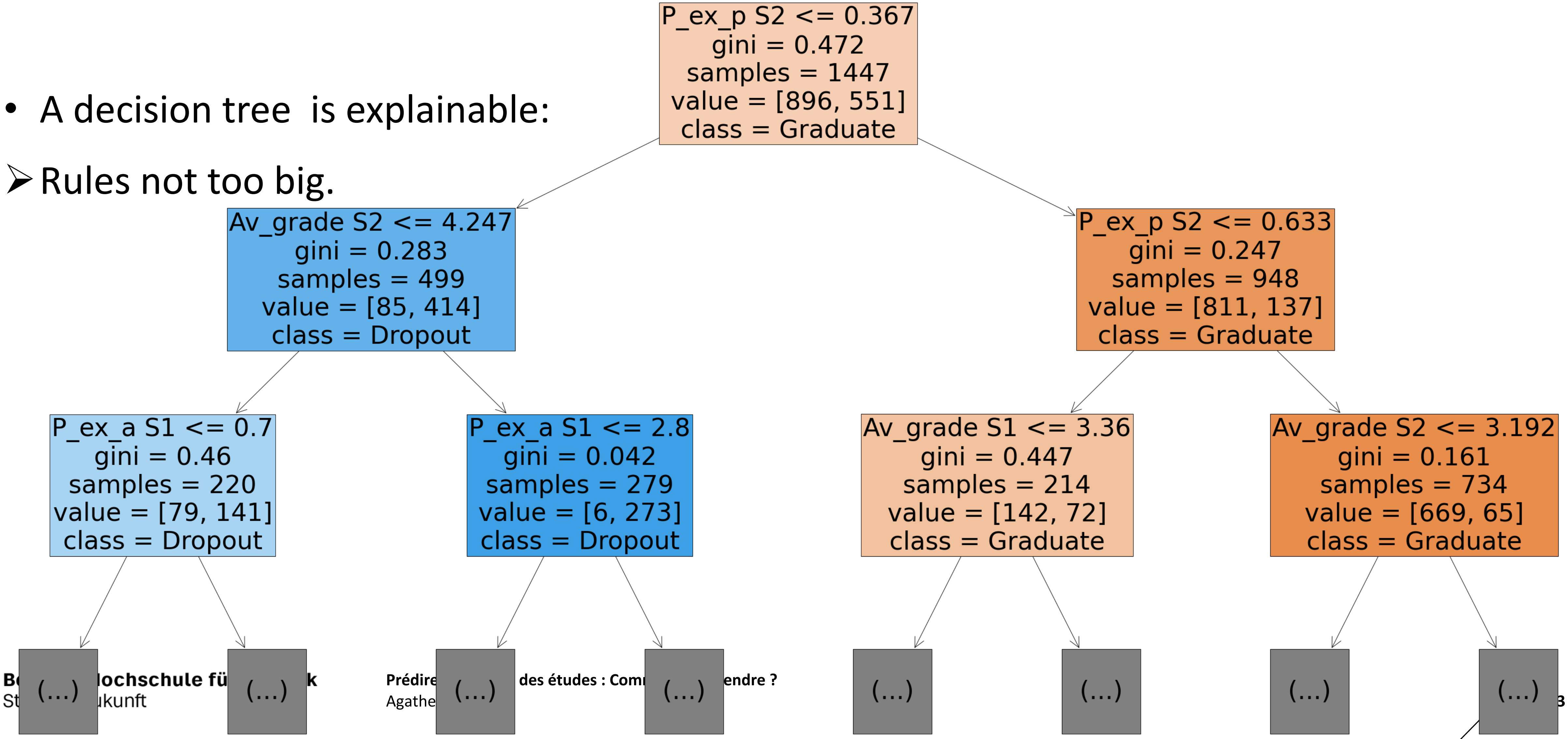


Model: Decision Tree

Predicting Dropout of a Degree: Decision Tree

- A decision tree is explainable:

➢ Rules not too big.



Predicting Dropout of a Degree: How good are the predictions?

Models built on the training set and evaluated on the test set.

		PREDICTED CLASS		
		Class=Yes	Class=No	
ACTUAL CLASS	Class=Yes	TP	FN	P
	Class=No	FP	TN	N

Predicting Dropout of a Degree: How good are the predictions?

Cross validation

- Partition data into k disjoint subsets
- k-fold: train on k-1 partitions, test on the remaining one

Predicting Dropout of a Degree: How good are the predictions?

- Evaluation:
  - Recall (Rec): from all students who dropped out, how many were found: TP/P?
  - Precision (Prec): from all students who were predicted to dropout, how many did really drop out: TP/(TP+FP)?
  - Accuray (Acc): Percentage of correct predictions (TP+TN) / (P+N).
  - Area under the curve (AUC): measures the confidence of the predictions; 0.5 means that the model does random predictions, 1 is the highest value.

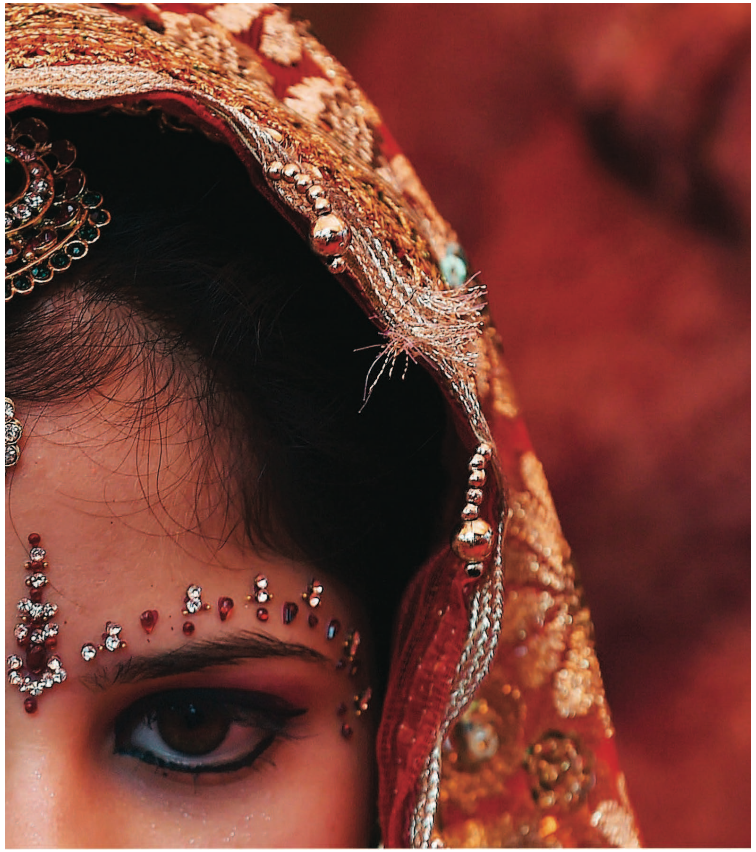
Predicting Dropout of a Degree

- No 100%!

Models		Prediction metrics				Set size	
Alg	Dataset	Rec	Acc	Prec	Auc	Train	Test
DT	1 All data	82.07%	**83.15%**	**92.79%**	**83.83%**	1447	362
	2 w/o 5%	**83.27%**	81.77%	89.70%	80.82%	1384	362
	3 w/o 10%	80.88%	81.22%	91.03%	81.43%	1312	362
	4 w/o 5% clusters 1-3	80.88%	81.49%	91.44%	81.88%	1432	362
	5 w/o 10% clusters 1-3	80.88%	81.77%	91.86%	82.33%	1428	362
LR	1 All data	79.28%	84.53%	98.03%	87.84%	1447	362
	2 w/o 5%	**80.48%**	**85.36%**	***98.06%***	***88.44%***	1384	362
	3 w/o 10%	80.08%	84.81%	97.57%	87.79%	1312	362
	4 w/o 5% clusters 1-3	79.28%	84.25%	97.55%	87.39%	1432	362
	5 w/o 10% clusters 1-3	79.68%	84.53%	97.56%	87.59%	1428	362
RF	1 All data	81.27%	85.08%	**96.68%**	**87.48%**	1447	362
	2 w/o 5%	82.87%	83.70%	92.86%	84.23%	1384	362
	3 w/o 10%	***84.86%***	***86.46%***	95.09%	**87.48%**	1312	362
	4 w/o 5% clusters 1-3	81.67%	84.53%	95.35%	86.33%	1432	362
	5 w/o 10% clusters 1-3	82.87%	85.36%	95.41%	86.93%	1428	362

Predicting Dropout of a Degree: Are the models fair?

Is it the picture of a bride? (Zou, J. & Schiebinger 2018)



Predicting Dropout of a Degree: Are the models fair?

What if the prediction Yes is related to something positive (“not defaulting on a loan”, “admission to a college”, “receiving a promotion” etc.) and the data used to train the model is skewed, like:

- The proportion of the applicants admitted for college is higher for white than for black students.
- The proportion of employees receiving a promotion is higher for males than for female employees.
- The proportion of female students dropping out is higher than the proportion of male students?

Predicting Dropout of a Degree: Are the models fair?

The model is likely to reproduce this bias:

- A white student might be predicted “admitted to college” with a higher probability than a black student.
- A male employee might be predicted “eligible for a promotion” with a higher probability than a female employee.
- A female student might be predicted “dropout” with a higher probability than a male student (higher recall).
  - How can we measure the fairness of models? Who is disadvantaged?

Predicting Dropout of a Degree: what for?

C2. Especially at the beginning of their studies, students often change their course of studies or drop out. Do you find a corresponding forecast helpful? Should explanations be provided as to how the system makes the forecast? What kind of explanations could be helpful for students to better manage their studies?

Student 2

S	P	Course	Grade
1	1	B01	3.3
1	1	B02	2.7
1	1	B03	3.0
1	1	B04	2.0
1	1	B05	5.0
2	1	B05	2.7
2	2	B06	5.0
2	2	B10	1.7
2	4/5	WP01	1.7

Student	Prediction	Probability
1	Graduate	89.07%
2	Dropout	72.35%
3	Graduate	91.30%

Example Student 2:
"The probability that you will successfully complete your studies is 27.65%."

C3. What support would you want in such a situation?

Predicting Dropout of a Degree: what for?

- Students’ answers: no clear trend.
  - Could be helpful –avoid studying too long- if presented with reasonable methods as people rarely like being told that they are bad/ struggling at something. Thoughtful formulation and effective support needed.
  - **Could demoralize some students**, could also reassure some students: self-fulfilling prophecy .
  - Students should be empowered to give feedback so that teaching could be improved and the model optimized.
  - Prediction has to be explainable.

Application: Support for Course Enrolment

- Which information do you typically use to decide which courses to take?
- What additional information would you like to have and why?

➢ Fellow Students' Decisions.

➢ Past Course Grades.

➢ Percentage of Passed.

➢ Friends and Acquaintances.

Application: Support for Course Enrolment

- To support struggling students, especially those who failed mandatory courses in their first semesters and are at risk of dropping out.
- Recommendations should be explainable.
- Recommendations should decrease the risk of dropping out.
- Recommendations should not *disturb* students who are doing well.

Application: Support for Course Enrolment

- Use enrolments patterns of student who graduated.
- Select the k-nearest neighbours of a specific student.
- Recommend courses that the majority of the neighbours has passed.

Application: Support for Course Enrolment

- Recommendations for student 0 based on three neighbours: M07, M08, M09, M15. Actually passed: M07, M15, M17, E06.

	S	1					2						3								
	C	M01	M02	M03	M04	M05	M05	M06	M07	M08	M10	E01	M07	M08	M09	M13	M14	M15	M16	M17	E06
#	ST																				
0	G	3.3	2.7	3.0	2.0	5.0	2.7	5.0			1.7	1.7	4.0	7.0	6.0	7.0	7.0	2.0	7.0	2.7	1.7
1	G	2.0	2.0	1.7	2.0	6.0	2.7	3.0		6.0	2.3	1.3		2.0	5.0	1.7		2.7		6.0	2.7
2	G	2.3	2.3	2.0	1.7	6.0	2.3	2.0		6.0	2.3	1.3	3.0	2.0	1.7						
3	G	2.3	4.0	2.0	2.0	6.0	3.3	4.0	6.0	3.7	2.0	1.7	2.0		3.3		5.0	2.3	6.0		

Application: Support for Course Enrolment

- First evaluation on historical data:
  - Match well the courses that student who graduate passed.
  - Recommends to enroll one course less to students who dropped out.

- Preliminary feedback of students:
  - Recommendations are explainable.
  - Suggestions for the user interface: marks should not be shown for instance.

Learnt Something new?

- Acquisition, stockage et utilisation de données éducationnelles administratives
- Algorithmes de prédiction
- Problèmes d’équité
- Utilisation possible des résultats

References

- Wagner, K., Hilliger, I., Merceron, A., Sauer, P.: Eliciting Students Needs and Concerns about a Novel Course Enrollment Support System. In Companion Proceedings of the 11th Learning Analytics and Knowledge Conference (LAK'21), p. 294-304. Workshop on Addressing Dropout Rates in Higher Education, Online - Everywhere, 2021.
- Novoseltseva, D., Wagner, K., Merceron, A., Sauer, P., Jessel, N., Sedes, F.:Wagner, K., Hilliger, I., Merceron, A., Sauer, P.: Investigating the Impact of Outliers on Dropout Prediction in Higher Education. In Proceedings of the Delfi Workshops 2021 at the 19th e-Learning Conference of the German Society for Computer Science, Dortmund-Online, Germany, September 13, 2021, p. 120-129.
- Wagner, K., Merceron, A., Sauer, P., Pinkwart, N.: Personalized and Explainable Course Recommendations for Students at Risk of Dropping out. To appear In Proceedings of the 15th International Conference on Educational Data Mining, EDM'2022, Durham, UK, July 24-27.

References

- Novoseltseva, D., Wagner, K., Merceron, A., Sauer, P., Jessel, N., Sedes, F.:Wagner, K., Hilliger, I., Merceron, A., Sauer, P.: Investigating the Impact of Outliers on Dropout Prediction in Higher Education. In Proceedings of the Delfi Workshops 2021 at the 19th e-Learning Conference of the German Society for Computer Science, Dortmund-Online, Germany, September 13, 2021, p. 120-129.
- Wagner, K., Merceron, A., Sauer, P., Pinkwart, N.: Personalized and Explainable Course Recommendations for Students at Risk of Dropping out. To appear In Proceedings of the 15th International Conference on Educational Data Mining, EDM'2022, Durham, UK, July 24-27.

MERCI !

Observations ?

Questions ?

BHT
Berliner Hochschule für Technik

Studiere Zukunft