A Tree-Based Framework for Student Performance Forecasting: Early Predictive Modelling and Individualized Behavioral Intervention

Filotas Theodosiou, Yves R. Sagaert, Liam Bossant, Tom Madou

Filotas Theodosiou

Contact Details:
- filotas.theodosiou@vives.be
- www.linkedin.com/in/filotas-theodosiou

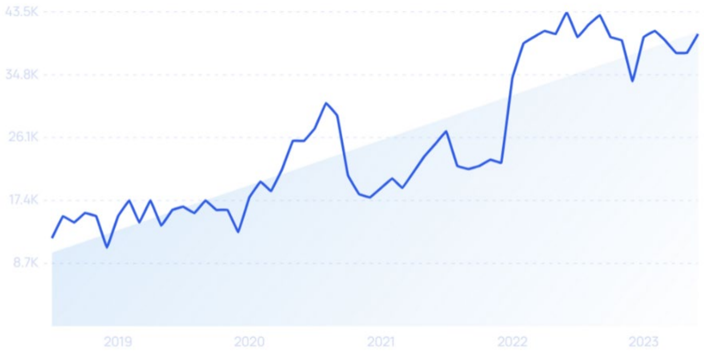
BELGIUM FORECASTING CENTRE

hogeschool vives

Agenda

A Brief Introduction to the Context

"The combined use of computer software and educational theory and practice to facilitate learning."



Searches for EdTech

An explosion in the volume of data available:

- Better understanding of the past
- More accurate future predictions

Learning Analytics In n a Nutshell

Data Collection | **Data Analysis** | **Actionable Insights**

Collect raw data from a learning platform

Apply descriptive and/or predictive modelling

Perform informed decisions based on the gained insights

Robinson, Rhonda; Molenda, Michael; Rezabek, Landra. "Facilitating Learning" (PDF). *Association for Educational Communications and Technology*. Archived (PDF) from the original on 22 September 2015. Retrieved 18 March 2016.

Learning Analytics: Benefits and Pitfalls

Reported Benefits of Learning Analytics:

- Reduce the dropout rate
- Assist in curricula design
- Create a more personalized learning environment
- Provide indications of students exam performance

The untold part of the story:

1. Accuracy over early predictions
Forecasting late leaves no time for learning adjustments

2. Teachers lack of confidence in complex predictive models
Limited faith in predictions they cannot interpret

Vikos Canyon in north-western Greece

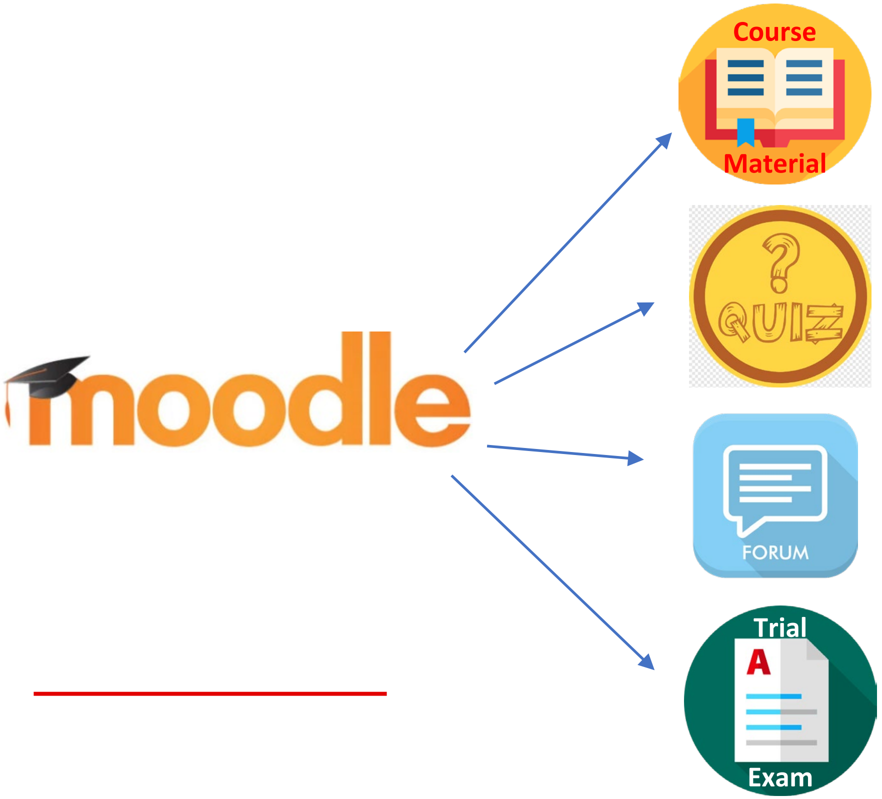
Cui, Y., Chen, F., Shiri, A.,& Fan, Y. (2019). Predictive analytic models of student success in higher education: A review of methodology. Information and Learning Sciences,120(3/4), 208–227. http://dx.doi.org/10.1108/ILS-10-2018-0104

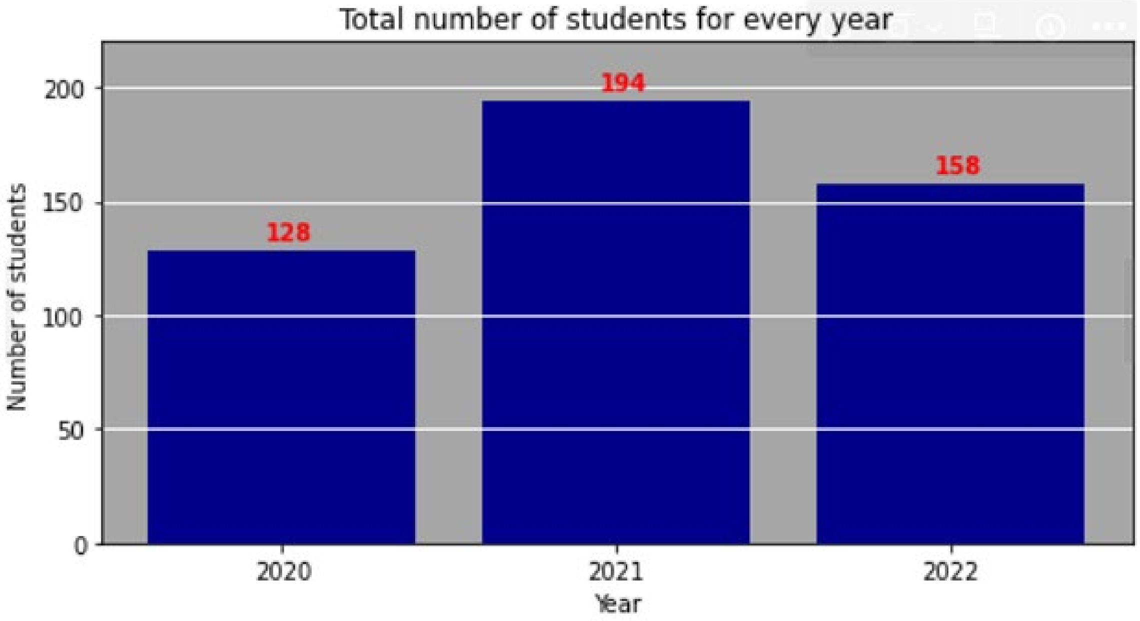
A Predictive Decision Support Framework📈

Setting up the Learning Environment

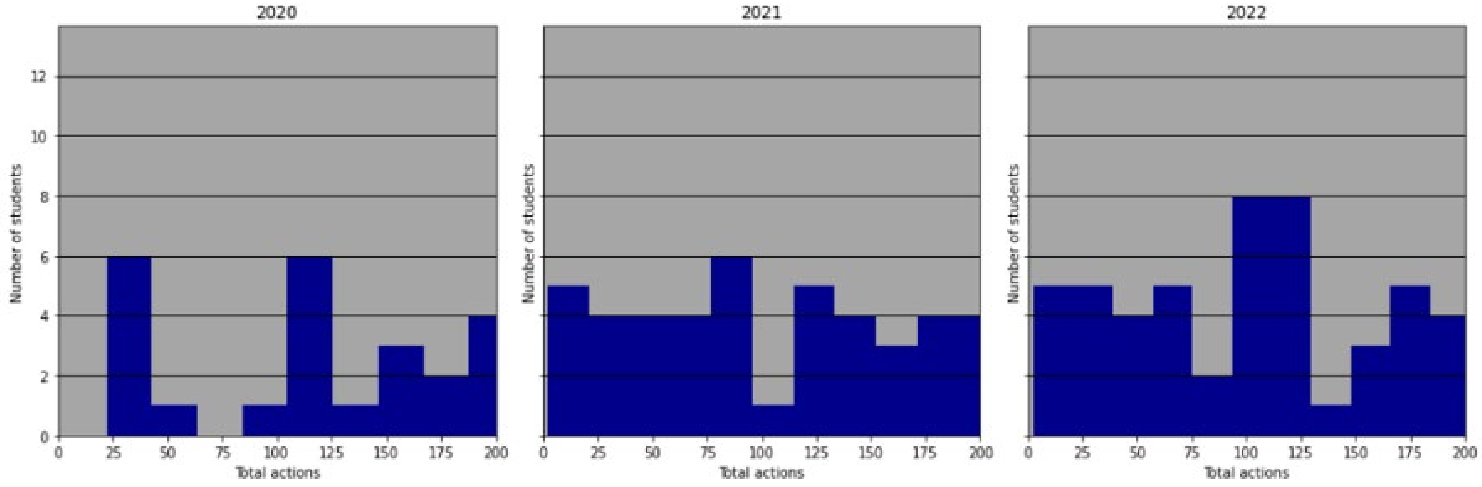
A single 1st year course.

- Data collected from 3 academic years
- Only the 1st semester is considered
- 2 years Remote Learning – 1 year Hybrid Learning





Potential issue with Data Shift?



An Interpretable Decision Support Framework:

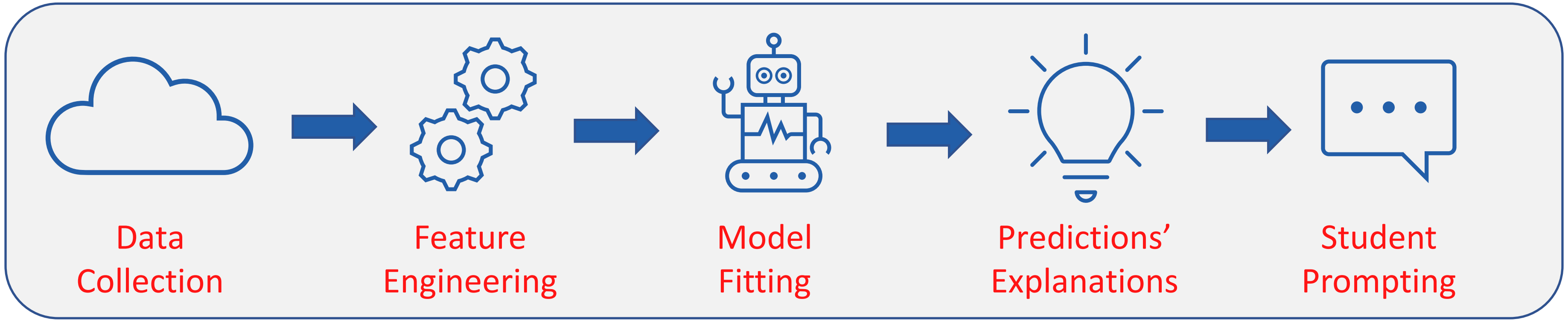
Using Predictive Modelling to Boost Students Exam Performance

The Objective of this Work: A Predictive Decision Support Framework

Out aim is:

- Predicting Early
- Predicting Accurately
- Predicting Interpretably

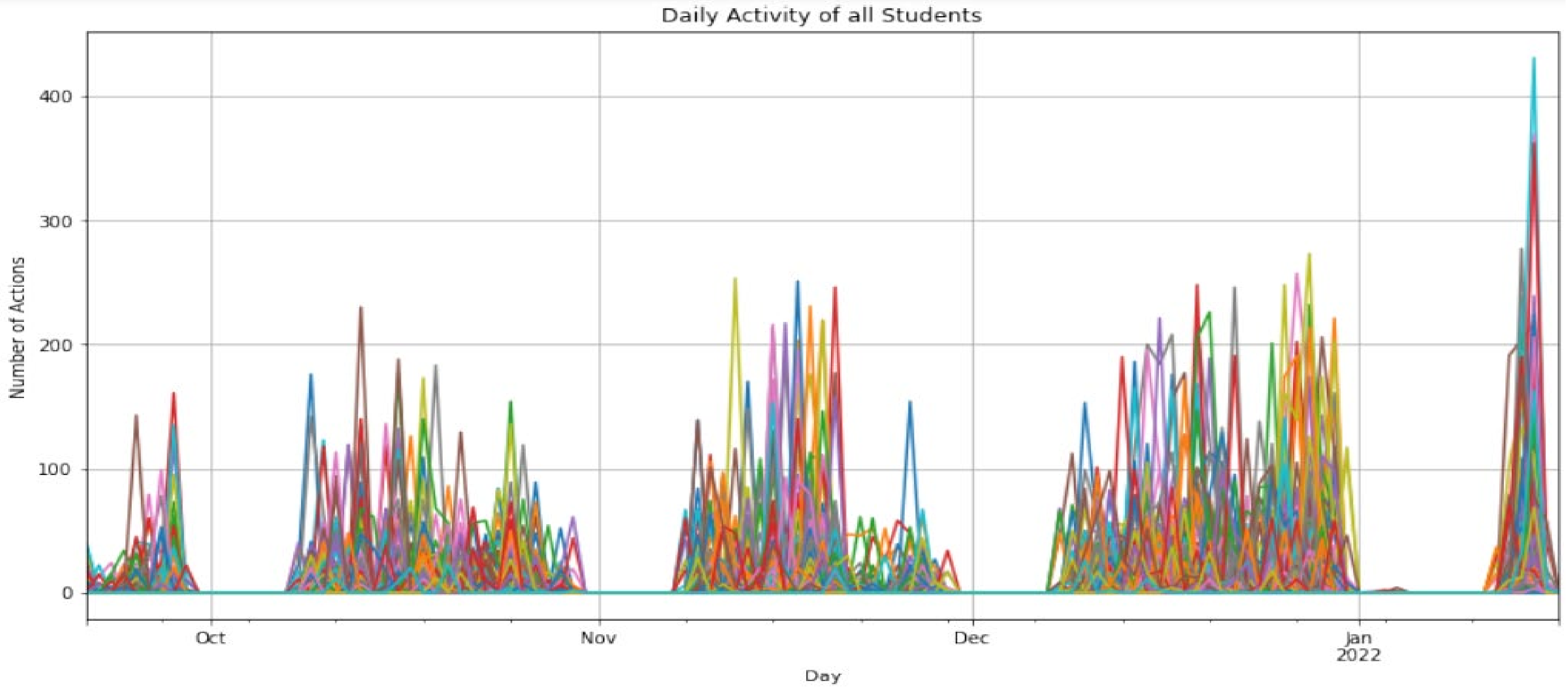
The Proposed Framework



Forecasting Methodology:
1. Predicting Early ⏰

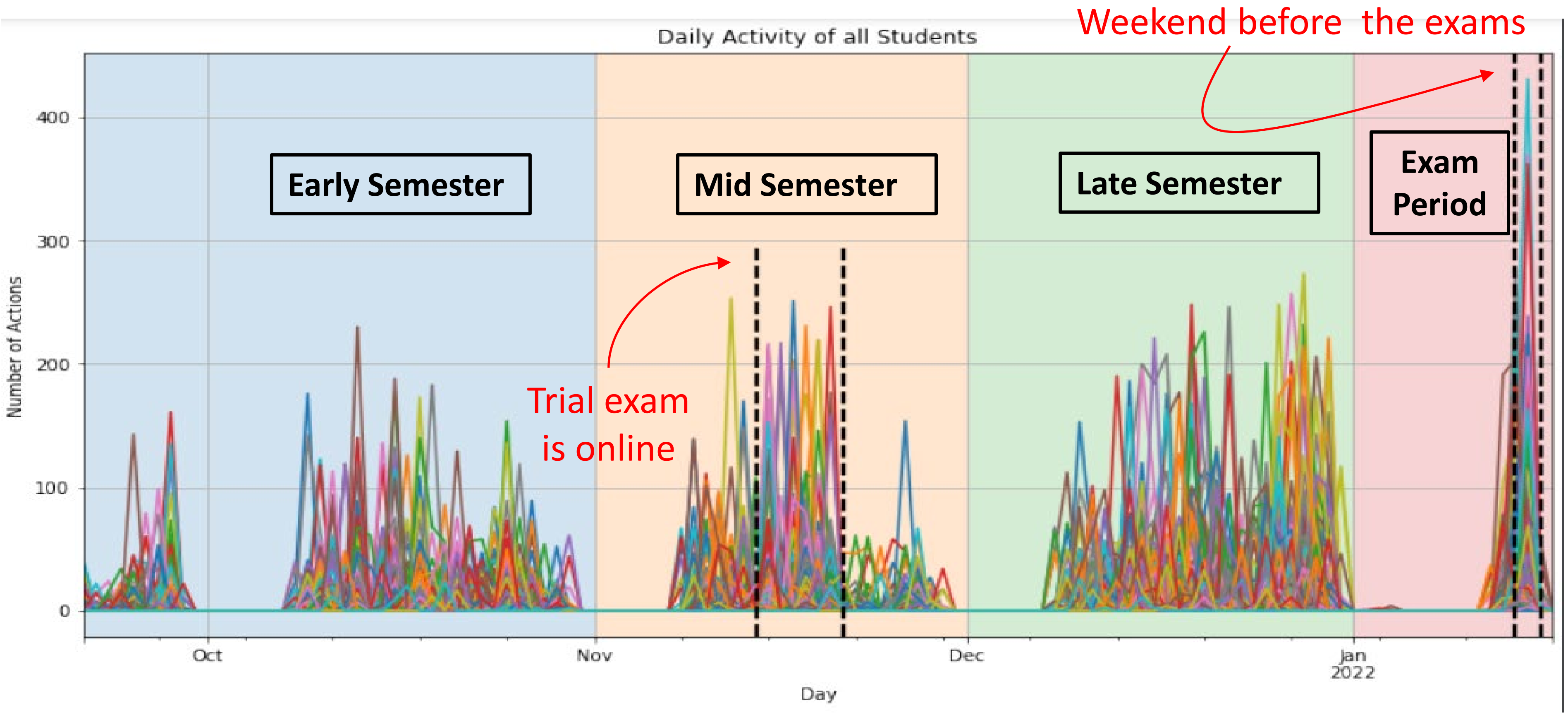
Forecasting Date Selection

- Leave Enough Time for Behavioral Adjustment



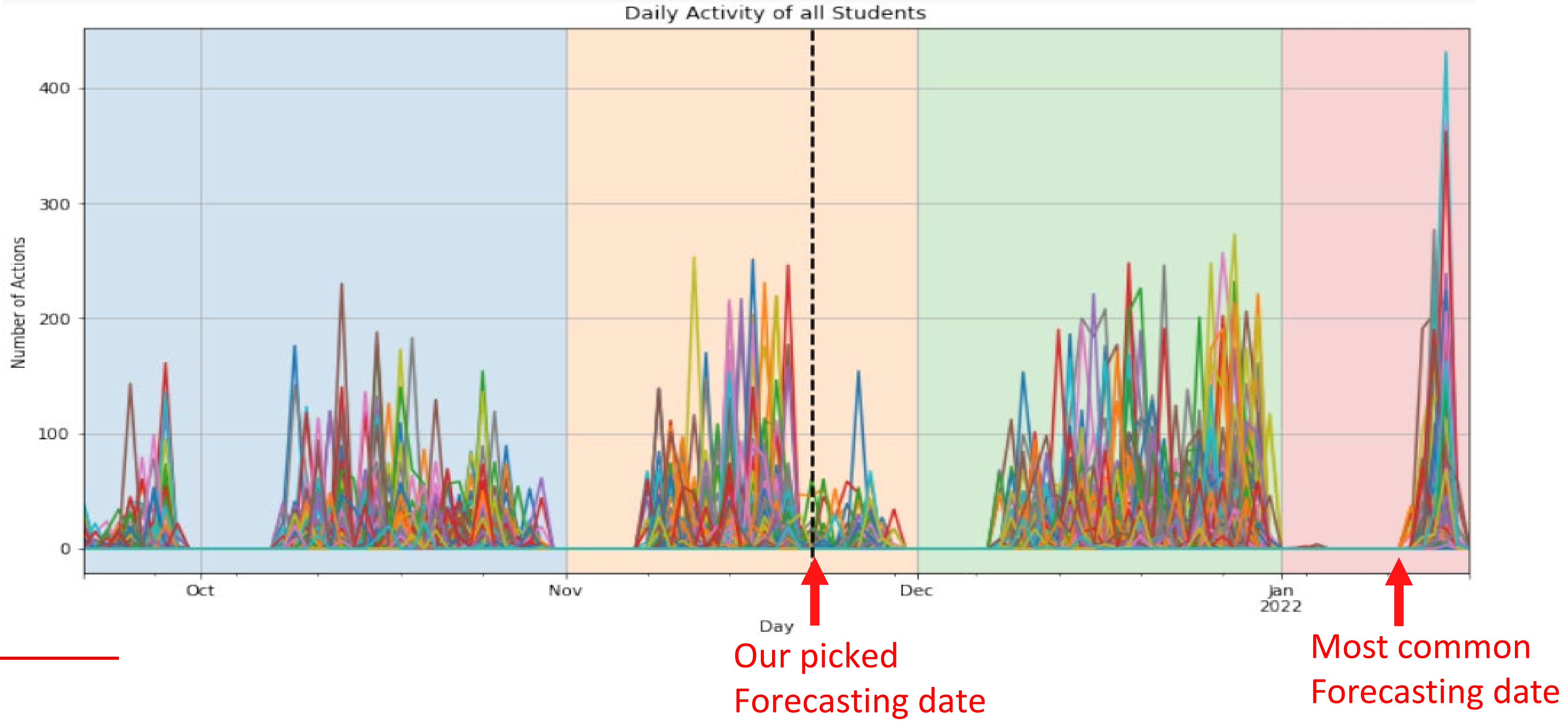
Forecasting Date Selection

- Leave Enough Time for Behavioral Adjustment



Forecasting Date Selection

- Leave Enough Time for Behavioral Adjustment



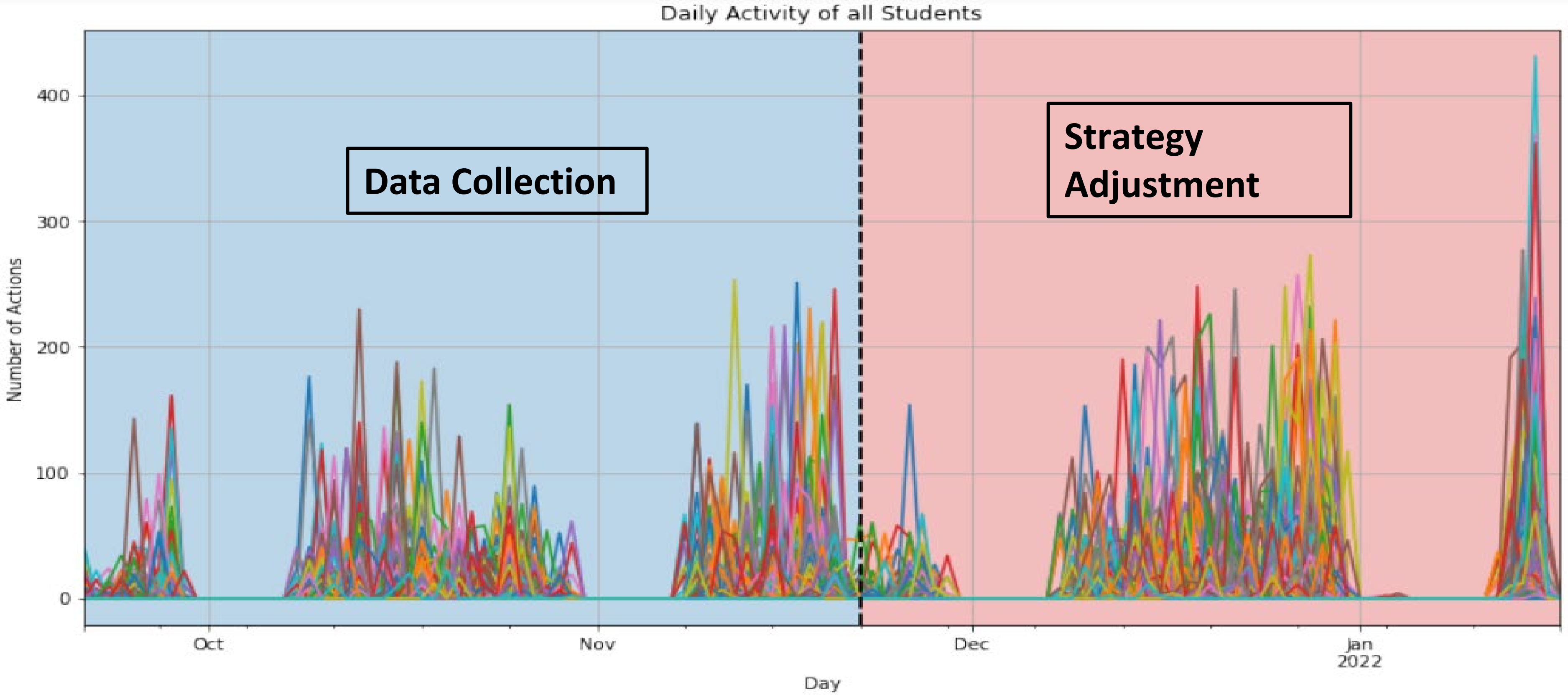
BELGIUM FORECASTING CENTRE

Forecasting Date Selection

- Leave Enough Time for Behavioral Adjustment

Collect Activity data for 55 days

Leave 61 days for learning adjustments



BELGIUM FORECASTING CENTRE

Forecasting Methodology:
2. Predicting Accurately 🎯

Model Selection: The thought process

Not entirely a forecasting task → Statistical methods not applicable.
We Rely on Data-Driven Machine Learning Methods.

Two prediction paradigms, one big debate.

Gradient Boosting Decision Trees: A good team of weak members

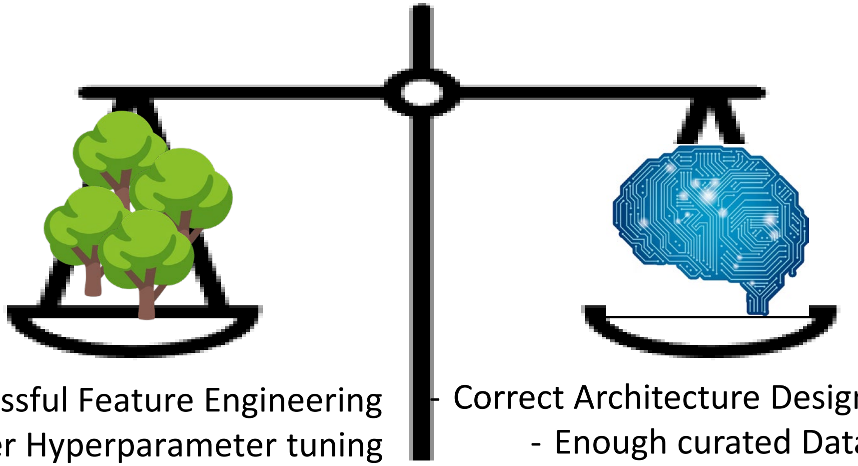
1. Different weak learners for diverse features
2. Informative feature space => more holistic view of the task
3. Requires detailed description of the problem

Neural Networks: A problem solver with unique pattern recognition skills

1. Multiple affine transformations
2. Finds meaningful representation of the feature space via deduction
3. Requires enough past experience and good architectural design

Which one to pick?

Conflicting literature findings



- Successful Feature Engineering
- Proper Hyperparameter tuning
- Skewed, longer tail feature distributions

- Correct Architecture Design
- Enough curated Data
- Regular and smooth feature distributions

McElfresh, Duncan, et al. "When Do Neural Nets Outperform Boosted Trees on Tabular Data?." *arXiv preprint arXiv:2305.02997* (2023).

Model Selection: Final Model Selection

Feature Engineering:

We construct multiple and advanced features from raw data.

Feature Set	Description
Regularity	Monitors Frequency of learning habits
Engagement	Monitors time spent on online course material
Performance	Monitors quiz and trial exam performance
Participation	Monitors participation in course related modules
Persistance	Monitors extent of task repetion
Student Profile	Personal characteristics and self-report learning traits

Model Selection

- We conduct an empirical investigation to identify the best model
- 5-fold cross validation on data of the first two years.

Bin students into 3 classes:

1. Grade < 8
2. Grade >8 & Grade < 13
3. Grade >13

Model	MSE	Recall	Accuracy
Logistic Regressio	-	0.38	0.4
XGBoost	10.83	0.68	0.53
CatBoost	**10.45**	**0.71**	**0.55**
LGBM	11.86	0.67	0.52
TabNet	14.58	0.65	0.42
FTTransformer	13.80	0.66	0.45

We pick CatBoost

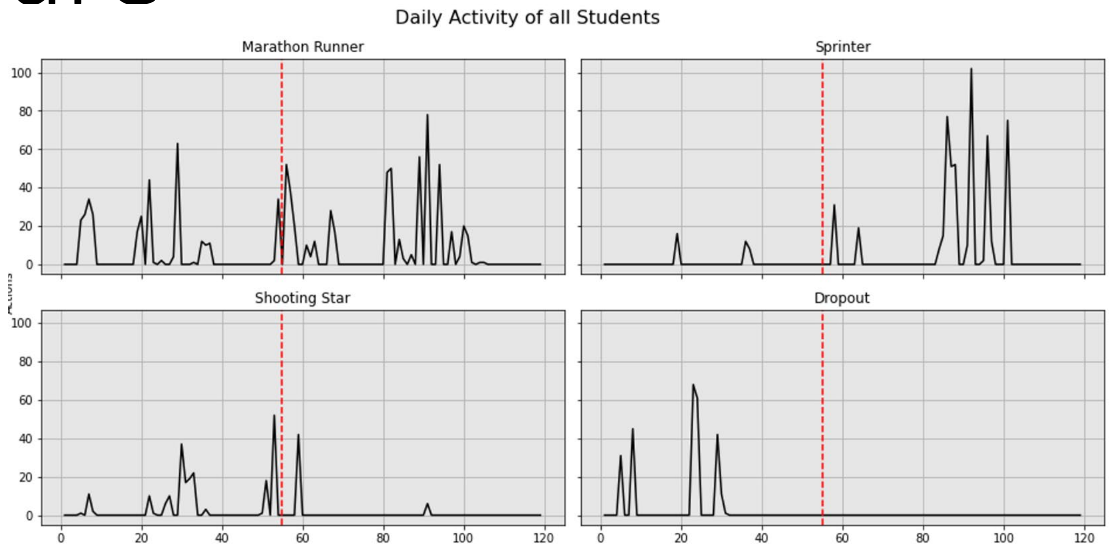
Its not just more accurate!

- Handles missing values
- Allows the usage of multiple features
- Compatible with explainability techniques

Embracing the Uncertain Future

How much faith can we put in a 1.5-month forecast of student performance?

- “Late Sprinters”, “Marathon Runners” and “Sugared-Up Sprinters”
- Unpredictable events: (eg break-ups, motivation adjustments)
- Pure luck?



Accuracy is just an illusion.

Enter Conformal Predictions (aka Empirical Methods)

Step 1: Estimate the conformal score: $\epsilon = f(y_{val} - \hat{y}_{val})$ on a validation set

Step 2: Estimate all quantiles q_a for $a \in (1{,}99)$ on ϵ

Step 3: For every prediction find q_a such as $\hat{y}_i \pm q_a \notin\ c_i\ \ wih\ c_i$ the binned class of $\hat{y}_i$

Step 4: Interpret q_a as the confidence in the prediction of $\hat{y}_i$

Forecasting Methodology:

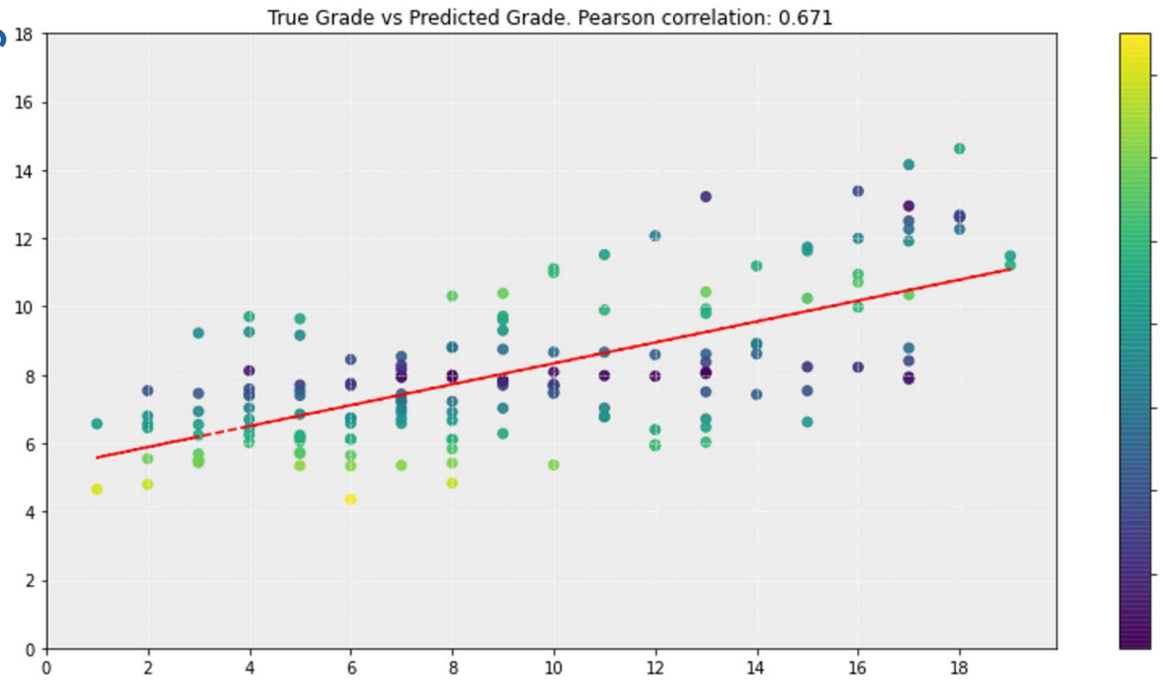
2. Predicting Interpretably

Explain your predictions

SHapley Additive exPlanations (SHAP) Values:

- Inspired by game theory
- Quantify the contribution of each feature to a single prediction
- Consider collaboration among features

A weighted sum of the marginal contribution of feature to a local prediction



SHAP values measure feature contributions to <u>predictions</u>, not true grades

Why SHAP Values you ask.

1. Identify which features contributed negatively to a student's predictions
2. Convert these features into areas of potential improvement for each student

Lundberg, Scott M., and Su-In Lee. "A unified approach to interpreting model predictions." *Advances in neural information processing systems* 30 (2017).

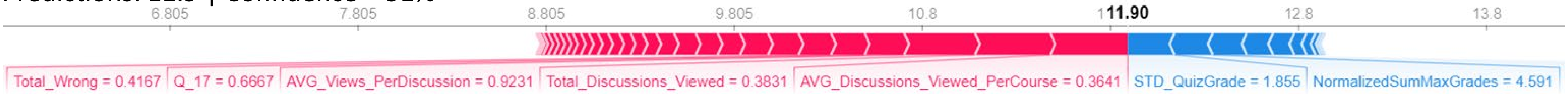
BELGIUM FORECASTING CENTRE

Some Explanations…

Student A:

Predictions: 11.9 | Confidence = 51%

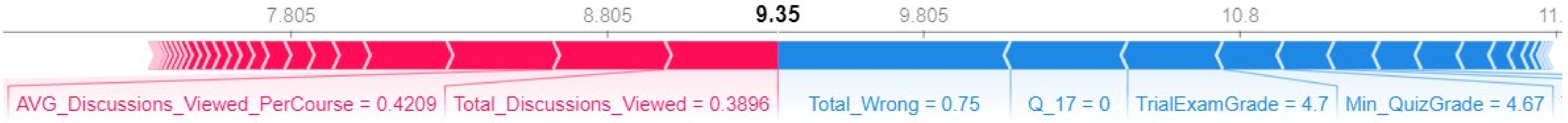
High Variance on Quiz Grades
Low Sum of Max Grades on Quizzes



Student B:

Predictions: 9.35 | Confidence = 82%

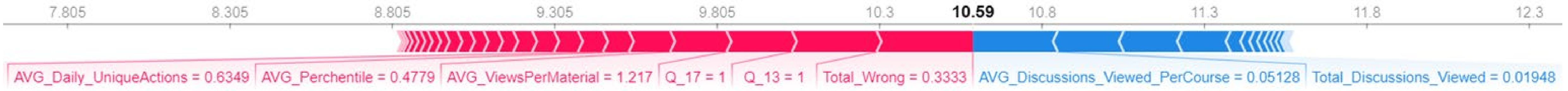
Bad Scores on Trial Exam
Low Minimum Score on a Quiz



Student C:

Predictions: 10.69 | Confidence = 67%

Low Forum Activity



Behavioral Adjustment

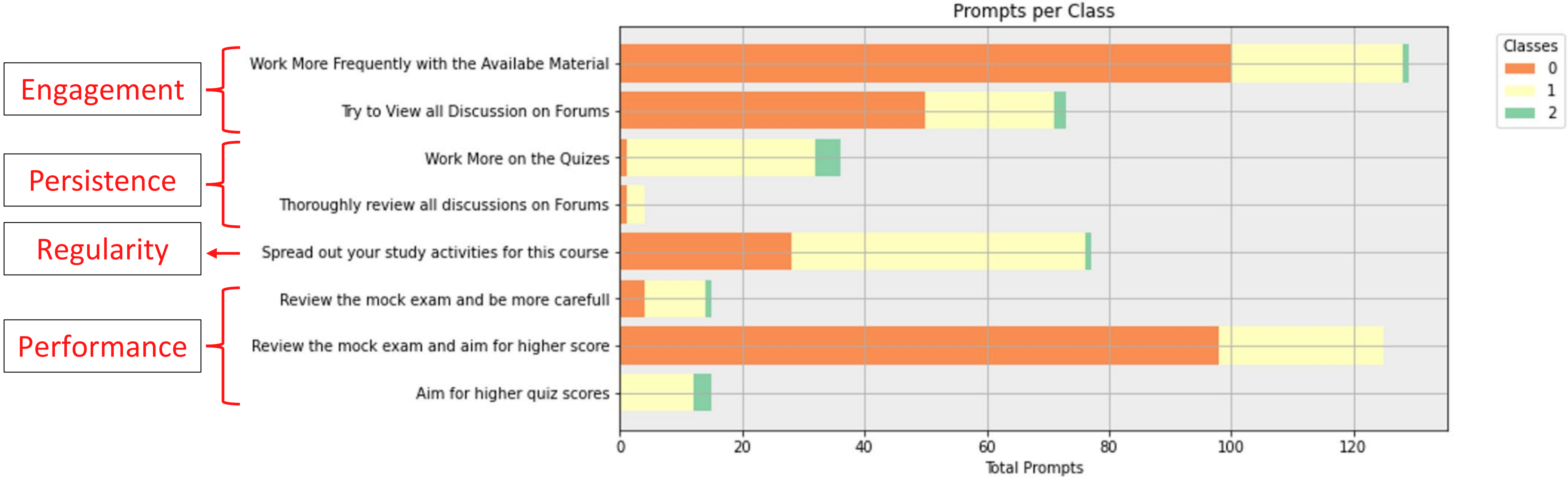
Turning Explanations into Prompts

Turn explanations into unique prompts:

Step 1: Match each feature with an appropriate prompt.
Step 2: Identify the 2 most negatively influencing features for every student
Step 3: Inform teacher for the prediction results and the prompt candidates
Step 4: Email students with the unique prompts.

Prompt Examples

Feature	Prompt
STD_Quiz_Grade	Work more on the quizes
STD_Daily_Actions	Work more frequently on the material
AVG_Discussion_per_Course	Thoroughly review all discussions on Forums
TrialExamScore	Review the Mock Exam and aim for higher scores

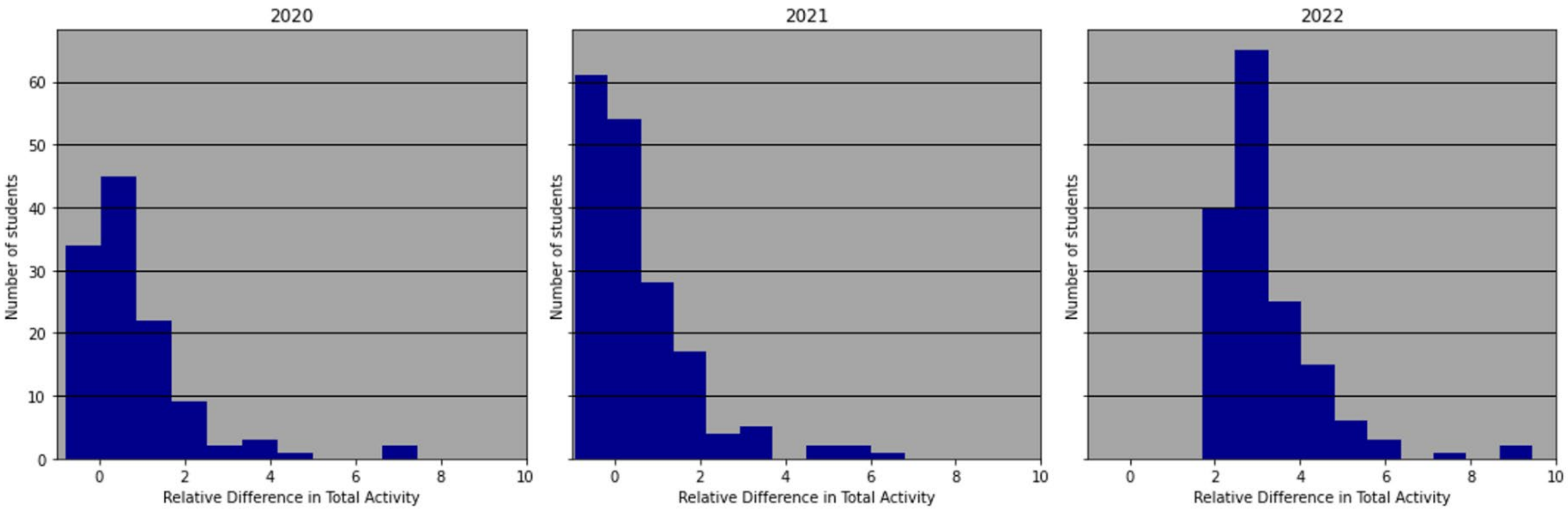


BELGIUM FORECASTING CENTRE

Did Prompting Worked? Some Initial Results

We define the Relative Difference in Total Activity as: $\log \frac{\frac{1}{n}\sum_{t=k+1}^{T} A_t - \frac{1}{n}\sum_{t=1}^{k} A\_t}{\frac{1}{n}\sum_{t=1}^{k} A\_t}$

In simple Terms:
How many more daily actions (on average) a student made after prompting



We conduct an ANOVA significance test

Comparisson	p-value	Significance
2022 vs 2020	2.66e-44	Yes
2022 vs 2021	3.51e-66	Yes
2022 vs 2020&2021	3.97e-78	Yes

BELGIUM FORECASTING CENTRE

Things to Discuss

1. We know prompting lead to learning adjustments.
Did it result in improvements in the exam performance?
We are investigating..

2. Can a fine tuned Large Language Model (LLM) become a great individual prompter?
Future Work: A personalized prompting assistant for every student

- Fine tune a LLM using Reinforcement Learning with Human Feedback (RLHF)
- Provide it with collected & predicted information for every student
- Collaborate with students.

To summarize:

We propose a **Predictive Decision Support Framework** for Learning Management Systems

We achieve:

- Early Predictions
- Accurate Predictions
- Transparent Predictions

Based on these predictions:

- We implement **specialized prompting**
- Initial results show **positive learning behavior adjustments**

In the Future:

We look to improve our prompting strategy using specialized AI agents for every student

The end ☺

Questions?

Filotas Theodosiou

Contact Details:
- filotas.Theodosiou@vives.be
- www.linkedin.com/in/filotas-theodosiou