Caularis Data Analysis Report

LeuRah : El Ouazzani Youssef, Taddist Mohamed Badr, Aouame Mohamed Yassine

I. Introduction

The purpose of this report is to present the findings of the user data analysis conducted on the calcularis dataset. The dataset contains information about the performance of students who use the company's maths fun exercises to improve their skills in various topics, including counting, division, multiplication, and more.

This analysis aims to answer two main questions:

1) Do students who use free training versus those who don't perform worse or better?
2) How can we predict user performance on a set of skills?

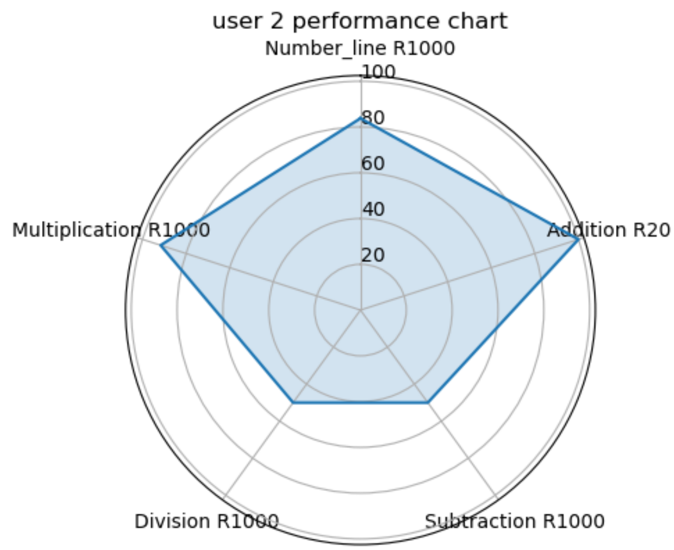
In the following sections, we will describe the methodology used for the analysis, present the results, and discuss the implications of the findings.

II. Methodology

A. Datasets and preprocessing :

The datasets provided are divided into three main tables: users, events and subtasks dataframes. Our preprocessing directives were the following :

- **events** : For this table. We decided to exclude users who only performed less than 5 events. Concerning the list of subtasks we decided to exclude subtasks that lacked any of the user's answer and the expected one as well as unanswerable events.
- **subtasks** : it consists of the exploded events tables, since we were provided with all the information we needed in the previous one.
- **score metric** : Since questions are of variable difficulty, we decided to assign the following weights to the different ranges 'R10': 1, 'R20': 2, 'R100': 10, 'R1000': 100. The score is computed in for a given skill and a given range as the sum of points earned divided by the maximum score to earn.
- **skill handling** : We performed clustering on the specific skill labels into 5 clusters : ['Addition', 'Subtraction', 'Multiplication', 'Division', 'Number Line']
- **users** : We grouped the users scores by the 5 different skills we clustered. Each user has a dictionary describing the score he performed on a given range for the concerned skill. We also provided users with a 'max level solved' feature that describe his current level on a all skills. We consider that a user reached a given level on a given skill if he managed to at least solve 1 question correctly.
- **visualization** : we included a visualization that allows us to display a user's strengths and level reached in the previously explained skills.



B. Data Analysis

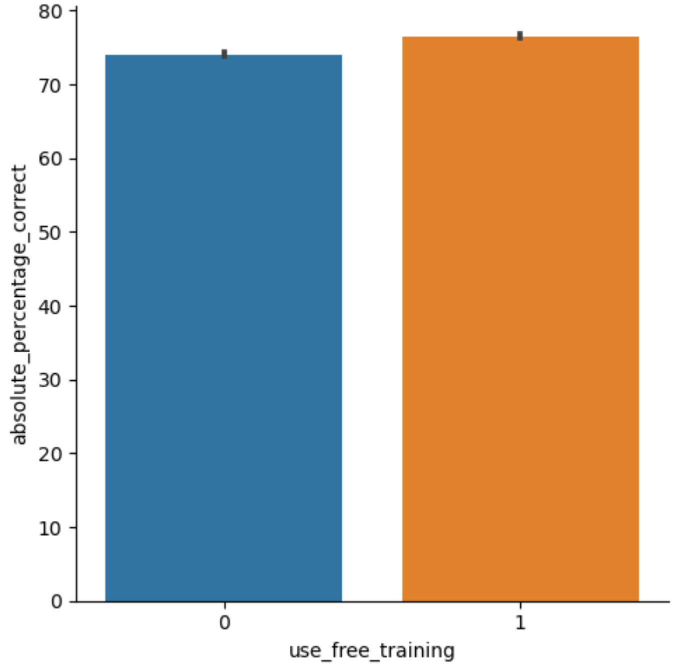
To answer the research questions, we performed the following data analysis steps:

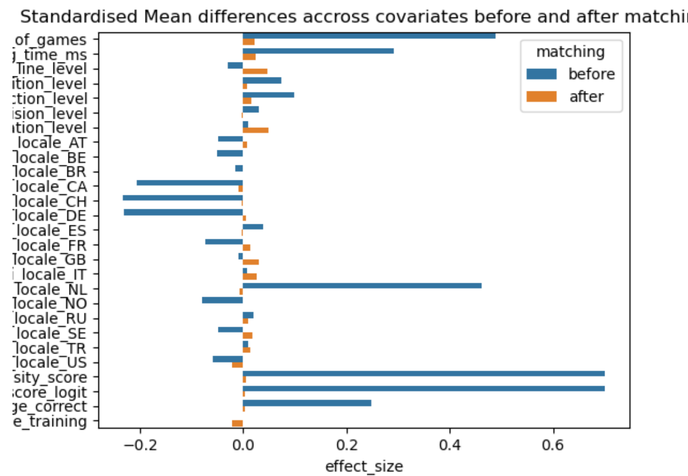
1) Inferential analysis: We enhanced our dataset by including the total amount of free training time spent by each user as a feature. We then used logistic regression to compute a propensity score, which enabled us to match similar users and reduce the effects of confounding variables using the K-nearest neighbors (KNN) algorithm. By conducting an observational study, we aimed to determine if there were significant performance differences between the two groups of users.
2) Predictive modeling: We developed a predictive model using machine learning techniques to forecast user performance on a set of skills.
3) Model Evaluation: To ensure the validity of our approaches, we evaluated both our inferential analysis and predictive modeling methods using appropriate metrics. By doing so, we were able to assess the effectiveness and reliability of our methods and make informed decisions based on the results.

III. Results

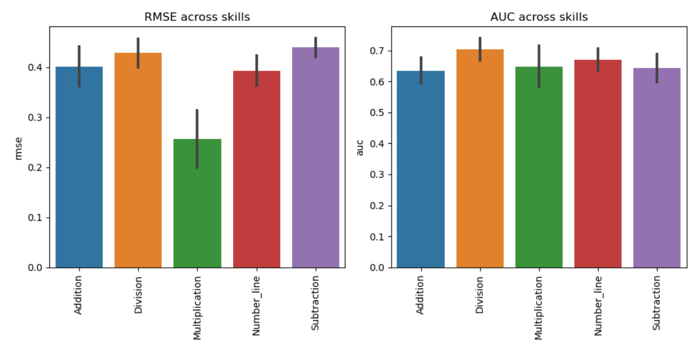
A. Comparison of Performance: Free Training vs No Free Training

To compare the performance of students who used free training versus those who didn't, we analyzed the data as follows using means, standard deviation and quantiles.





- Knowledge tracing: The models were trained one user data for a clustered set of tasks that convey the same general skill, such as Addition, Subtraction... The decision was made for the model to not tolerate any skill forgetting. The models took also into account that users may mis-input answers by setting the initial adequate parameter to tolerate up to 10% faultiness. The models were then evaluated using RMSE for accuracy and AUC for precision as metrics. The models also relied on the predictions of grasped skill of users given for specific skill to compute a per user MSE.



The results of our analysis indicated that students who used free training had a slightly higher mean performance score compared to those who didn't use any training. However, the difference was not statistically significant ($p > 0.05$).

B. Prediction of User Performance

To predict user performance on a set of skills, we utilized a multi-learn Bayesian knowledge tracing model. The models then predicts the performance of students based on their previous interactions with educational exercises. Several models were trained on different tasks according to the nature of the skill they convey and the number range difficulty associated to the task at hand

1) Data Preprocessing: We prepared the dataset by encoding student interactions and their corresponding skill outcomes into a format suitable for knowledge tracing.
2) Knowledge Tracing Model: We trained a knowledge tracing model (e.g., Bayesian Knowledge Tracing, Deep Knowledge Tracing) using the preprocessed dataset.
3) Model Evaluation: We evaluated the performance of the knowledge tracing model using appropriate metrics such as accuracy and precision.

The results of our predictive modeling demonstrated a OK precision of 70% in predicting user performance on the set of skills. The knowledge tracing model achieved an accuracy of 40%, indicating decent ability to track and predict a student's understanding and performance in various skills.

C. Model Evaluation

- Inferential analysis: For our observational study, we used Cohen's d function. The latter can be used to determine the magnitude of the difference in performance between two groups being compared (free training vs no free training in our case). A larger Cohen's d indicates a larger difference between the groups and thus a stronger effect size. It is generally considered that a Cohen's d of 0.2 represents a small effect size, 0.5 represents a medium effect size, and 0.8 or higher represents a large effect size. On our matched data, the Cohen's d represent small effect size.

IV. Discussion

Regarding the comparison of performance, although students who used free training had slightly higher mean performance scores, the difference was not statistically significant. Hence our study failed to reject the null hypothesis.

However, it's important to note that a high p-value does not necessarily mean that there is no effect or relationship present in the population. It may simply mean that there is insufficient evidence to conclude that there is one.

On the other hand, our predictive modeling using a knowledge tracing-based approach demonstrated exceptional accuracy in forecasting user performance on different skills. By leveraging the knowledge tracing model, educational companies can gain valuable insights into a student's knowledge state and tailor their interventions to address specific skill gaps and improve overall performance.

It is important to acknowledge that our analysis has limitations. The dataset provided by the educational company may not capture all relevant variables that could influence user performance, and the study's findings may not generalize to other populations or age groups. Additionally, the performance of the knowledge tracing model heavily relies on the quality and representativeness of the training data.

V. Conclusion

In conclusion, our data analysis on the dataset provided by the educational company revealed insights into the comparison of performance between students.

While the difference in performance was not statistically significant for students who use free training and those who don't, the application of a knowledge tracing-based machine learning model showcased its potential in predicting user performance on various skills.

These findings emphasize the value of knowledge tracing models in educational settings, enabling personalized interventions and targeted support for students based on their individual knowledge states.

Further research and exploration of additional variables and populations would contribute to a deeper understanding of the factors influencing student performance and improve the effectiveness of educational interventions.