Energy Consumption of AI in Education: A Case Study

Vorname1 Nachname 1[1], Vorname2 Nachname2[2] und weitere Autorinnen und Autoren in der gleichen Notation

Abstract: Although the utilization of AI in education has grown considerably in the last decade, its environmental impact has been disregarded thus far. In this paper we examine the energy consumption of Artificial Intelligence (AI) in education, which is employed, for instance, in adaptive learning. We measured the energy requirements of four AI implementations used on the student learning platform XX. We found that two of the implementations have notably low energy and CPU demands in comparison to the baseline, while in two other implementation these parameters are significantly higher. We conclude that more attention should be paid to whether comparable performance of AI in education can be achieved with lower energy consumption.

Keywords: Green IT, Artificial Intelligence in Education, Sustainable Learning Analytics

1 Introduction

The significance of sustainability within the Information and Communication Technology (ICT) industry has increased in recent years. The sector was responsible for 9% of the world's energy consumption in 2018, and this figure is estimated to rise to 20% by 2025 [MGC21]. The reason for this increase is the persistent expansion of the ICT industry which outpaces growth in other sectors [BE18]. Therefore, the amount of processed data as well as the number of (intelligent) applications grow and result in more energy that is needed to run these. For instance, the computing power to train the largest deep learning models has increased by the factor 300,000 in six years [Be21].

The number of applications using machine learning in education is rising sharply as well [RV20]. To this point, only the learning benefits have been taken into consideration, which include variables e.g., learning success, dropout rate etc. However, applications supported by machine learning require more energy than standard digital learning systems. Although many studies have recently looked at the energy consumption of application of AI [SGM19], we did not find any that covered the use of energy of AI in education. We aim to fill this gap by measuring the use of energy in prediction models that are applied on the AI-based adaptive learning platform XX by using a measurement setup as proposed by [Wo22], which is described in detail in section 3. We evaluate different options of transforming a digital learning platform into an adaptive learning platform by using machine learning model and interventions. The reflections concerning energy

[1] Einrichtung/Universität, Abteilung, Anschrift, Postleitzahl Ort, emailadresse@author1, https://orcid.org/0000-0000-0000-0000

[2] Einrichtung/Universität, Abteilung, Anschrift, Postleitzahl Ort, emailadresse@author2

consumption are secondary to the learning impact; however, it is crucial to discuss the trade-off between learning benefit and use of energy. Given the same learning benefit, the lower energy consuming technology should be preferred.

Therefore, our research question is: By how much does the local resource utilization differ among different AI models?

The structure of this paper is organized into several sections. Firstly, we will present some background information. After, we will describe the platform and test setup, followed by an analysis of our results and a discussion.

2 Background

Dick et al. define sustainable software as a “software whose direct and indirect negative impact on economy, society, human beings, and the environment resulting from development, deployment, and usage of the software is minimal and/or has a positive effect on sustainable development” [DNK].

When measuring the actual consumption of energy in a software, one can use a software-based or hardware-based approach. A software-based approach uses software tools to measure the energy consumption while a software is executed. Since the consumption is only estimated, inaccuracies might occur. On the other hand, a hardware-based measurement uses physical energy meters that are connected to PC hardware. This approach is more expensive but provides more accurate results [MGC21]

An adaptive learning environment is defined as a learning environment that uses AI to adapt to a user’s unique level of skills in real-time to support the learning process using educational data [Me19].

3 Methodology

XX, the platform we analyzed, is a learning platform for German grammar and spelling exercises, addressing students from 5th grade onwards which has been established in 2008. Ever since, the platform has been accessed by more than 1.2 million users and more than 12 million tasks have been completed. Elements of gamification are used and tasks that were answered incorrectly are asked again (XX). Four different AI models (decision tree, logistic regression, support vector machine (SVM) and multilayer perceptron (MLP) have been trained on datasets that have been recorded in 2020 and have been applied in an experiment on the platform for a four-months period to predict the probability of solving a task correctly (XX). The models performed differently with regards to accuracy and precision: The decision tree’s accuracy was 97.04% and 97.75% precision; the logistic regression model returned a 96.94% accuracy and a 97.76% precision while it was

96.62%/ 97.38% in the SVM and 97.09%/ 97.97% in the MLP. Therefore, MLP and decision tree outperform the other models.

To measure the energy consumption accurately, we used a measurement protocol described by [Wo22]. Before every measurement, the computer is rebooted to minimize the number of programs running in the background or using local storage. The setup measurement consists of a System under Test and energy data aggregator are started as well as the Workload Generator, resource management and energy data collection. In addition, a Network Time Protocol server is set up to provide an accurate time for the measurement environment and finally assigns the measurement data in the evaluation, while the energy data aggregator queries and logs the energy data via Simple Network Management Protocol. In order to analyze the results, the OSCAR tool (Open Source Software Consumption Analysis and Reporting) is being accessed by the control and evaluation station. To avoid measurement errors and inaccuracies due to user interaction, a script is run to perform tasks and measurements.

Because of the model requirements, we needed to install Python in two different versions: version 3.7 for decision tree, logistic regression and SVM and version 3.10 for the MLP. Since calling the SVM requires more time than other models, we decided to use a separate measuring time for this model (600 seconds for the model and 60 seconds for cool down), while all other models were measured in 120 seconds and 60 seconds cool down. The measurements are run 30 times and the average is calculated over these values [Wo22]. We compared our measurements to baseline measurements (new set up computer with nothing but windows installed) and idle mode (having all software requirements for the measurements installed) measurements. We utilized a t-test as a statistical tool to analyze our results in comparison to the baseline.

4 Results

We found differences to baseline and idle with regards to CPU intensity and consumption of electric energy in Watt hours (Wh) and power in Watt.

In table 1, the results for baseline and idle measurements are presented. One can see that the differences among the parameters displayed are minor which suggests that the installed software does not cause a large additional load.

	BL 120 + 60	Idle 120 + 60	BL 600 + 60	Idle 600 + 60
Mean el. power in W	32,47	32,43	32,29	32,29
Mean el. work in Wh	1,08	1,09	5,38	5,38
Average CPU load	2,26%	1,71%	1,33%	1,08%

Tab. 1: Baseline and Idle Measurements (BL: Baseline)

Table 2 presents the results of our measurement runs. The average power measurements of decision tree, logistic regression and MLP was between 32.32W and 33.48W, showing only a slight change to the baseline value of 32.47W and the idle time measurement of 32.43W. On the other hand, the SVM measurement has a power demand of 42.4W, which is a significant increase to the measurement of 32.39W in both the baseline and idle run. We can also see that the short running measurements have a low electric work in Watt hours (Wh) compared to baseline and idle measurements, only the MLP has slightly higher values. On the other hand, the SVM measurement resulted in relatively higher values as the electric work has increased by 156,04% in comparison to the baseline and 154,61% in comparison to the idle measurement.

Looking at the CPU utilization, we can see that it moves between 2.32% in the decision tree and 2.37% in the logistic regression, showing only a slight increase to baseline and idle measurements that is not statistically significant. We found a CPU load of 5.5% in the MLP while the SVM had a higher average CPU usage of 13.14%, making this highly statistically significant compared to the baseline values.

	Measurement Results	t-Test	Delta % baseline	Delta % idle
Mean el. power in W DT	32,36	2,22 (*)	-0,34%	-0,22%
Mean el. work in Wh DT	1,09	-	0,93%	0,00%
Av. CPU load DT	2,32%	-0,05 (n.s.)	2,65%	35,67%
Mean el. power in W LR	32,32	1,16 (n.s.)	-0,44%	-0,32%
Mean el. work in Wh LR	1,09	-	0,93%	0,00%
Av. CPU load LR	2,37%	-0,54 (n.s.)	4,87%	38,60%
Mean el. power in W MLP	33,48	-11,41 (***)	3,11%	3,23%
Mean el. work in Wh MLP	1,15	-	6,48%	5,50%
Av. CPU load MLP	5,50%	-15,47 (***)	143,36%	221,64%

	Measurement Results	t-Test	Delta % baseline	Delta % idle
Mean el. power in W SVM	42,40	-135,15 (***)	31,14%	31,18%
Mean el. work in Wh SVM	7,07	-	156,04%	154,61%
Av. CPU load SVM	13,14%	-135,12 (***)	522,44%	705,54%

Tab. 2: Measurement Results (DT: Decision Tree, LR: Logistic Regression)

5 Discussion

Our results show differences between the models, suggesting that the choice of a specific model does have an impact on local energy and system requirements. While decision tree and logistic regression hardly use more energy than the baseline, we found that the increase of using energy and CPU utilization in the MLP and SVM is highly significant. This correlates with the general findings of [SGM19], who have looked at the energy consumption of training and developing different Natural Language Processing algorithms and found that power use and training time differs across models. However, our measurements have been applied to different models than these researchers used and looked at implementations of AI in education in particular.

We can only measure the local energy consumption since we do not have access to data centers. We only consider the energy that is consumed while running the software and do not take manufacturing emissions into account. Also, having used different Python versions might influence the results.

We ran the measurements in a standard setup on a reset computer using 30 iterations including cool down time to get the best possible results. Nevertheless, we cannot exclude measurement errors due to unforeseen background activities. As this is a standard setup that does not apply to every user of this application, we assume that the accurate values will vary among users. Nevertheless, they provide strong implications on which data model is rather resource efficient or intensive.

To verify the results, we suggest rerunning the experiment with either a different dataset and/or a different measurement setup.

6 Conclusion

Our study suggests that decision tree and logistic regression models in an implementation of AI in education have the lowest impact in terms of energy and system requirements.

Although the differences seem rather small, they do have an impact in an application frequently used by a large number of people. To address the ICT sector's emissions reduction requirements and contribute to climate change mitigation efforts, implementing energy-conscious AI technology should be a priority on educational platforms. However, uncertainties remain. Due to the scarcity of previous works in this area, this case study should be rerun to confirm the results.

Overall, we could see that the choice of a specific algorithm has an impact on the resources consumed. Therefore, we encourage developers to not only consider accuracy and performance, but also the environmental impact of a software.

References

[BE18] Belkhir, L.; Elmeligi, A.: Assessing ICT global emissions footprint: Trends to 2040 & recommendations. Journal of cleaner production 177, pp. 448–463, 2018.

[Be21] Bender, E. et al.: On the Dangers of Stochastic Parrots: Proceedings of the 2021 ACM 2021, pp. 610–623, 2021.

[DNK] Dick, M.; Naumann, S.; Kuhn, N.: A Model and Selected Instances of Green and Sustainable Software. In (Berleur, J.; Hercheui, M. D.; Hilty, L. M. Eds.): What Kind of Information Society? Governance, Virtuality, Surveillance, Sustainability, Resilience. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 248–259, 2010.

[Me19] Meier, C.: KI-basierte, adaptive Lernumgebungen, 2019.

[MGC21] Mancebo, J.; García, F.; Calero, C.: A process for analysing the energy efficiency of software. Information and Software Technology 134, p. 106560, 2021.

[RV20] Romero, C.; Ventura, S.: Educational data mining and learning analytics: An updated survey. WIREs Data Mining and Knowledge Discovery 3/10, 2020.

[SGM19] Strubell, E.; Ganesh, A.; McCallum, A.: Energy and Policy Considerations for Deep Learning in NLP. arXiv, 2019.

[Wo22] Wohlgemuth, V. et al. Eds.: EnviroInfo 2022. Short-/work in progress-papers September 26-28, 2022, Hamburg, Germany. Gesellschaft für Informatik e.V. (GI), Bonn, 2022.