Sustainable Learning Analytics: Measuring and Understanding the Drivers of Energy Consumption of AI in Education

Marlene Bültemann[a], Katharina Simbeck[b], Nathalie Rzepka[c] and Yannick Kalff[d]
Hochschule für Technik und Wirtschaft Berlin, Treskowallee 8, 10318 Berlin, Germany

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

Abstract: As learning analytics increasingly relies on machine learning (ML) to provide insights and enhance educational outcomes, the environmental impact of these ML-driven tools has become a critical but underexamined issue. This study aims to fill this gap by investigating the energy consumption of various machine learning models commonly employed in learning analytics. This is by the execution of four distinct models — Support Vector Machines (SVM), Multi-Layer Perceptrons (MLP), Decision Trees (DT), and Logistic Regression (LogReg) — when applied to an educational data set. Our findings reveal significant disparities in energy consumption between these models, with SVM and MLP models consuming considerably more energy than their simpler counterparts. This research serves as a call for action for the learning analytics community to prioritize energy-efficient AI models, thereby contributing to broader sustainability goals in the face of climate change.

1 INTRODUCTION

Climate change is one of the most urgent issues of our time that threatens humans and the planet and requires immediate action (IPCC, 2022). Mitigation efforts are required in all sectors, including the information and communication technology (ICT) sector (Anser et al., 2021), which contributed to 9% of global emissions in 2018, estimated to rise to 20% until 2025 (Mancebo et al., 2021).

While there is growing awareness about hardware and data center emissions, the environmental impact of software products remains underappreciated. Software is intangible, and no visible waste is generated when used or disposed of (Naumann et al., 2021). The environmental costs of software, in general, remain a niche: "In many software development projects, sustainability is treated as an afterthought, as developers are driven by time-to-market pressure and are often not educated to apply sustainability-improving techniques" (Durdik et al., 2012). Nevertheless, "our civilization runs on software" (Stroustrup, 2014), which leads to an increased demand for non-sustainable products (Calero and Piattini, 2017).

Above that, applications utilizing artificial intelligence (AI) have risen significantly and will continue to do so (Wu et al., 2022). AI research is continuously improving the accuracy and capabilities of AI models, which requires growing computational power (Henderson et al., 2020; Schwartz et al., 2020).

The increased demand also includes AI for learning analytics (LA). While positive effects on learning success or reduced drop-out rates are well-documented benefits, resource consumption of LA remains a research gap. In a literature review on sustainability dimensions of e-learning, most of the examined studies have dealt with individual and social requirements (Alharthi et al., 2019). Only 4% of the publications dealt with environmental aspects and mainly focused on cloud computing. Despite the growing integration of AI in education, there has been no research on its resource consumption.

Our research contributes to this research gap and compares the energy consumption of different LA AI models. Our research questions are:

RQ 1: How do different AI models used in education vary in terms of energy efficiency and environmental impact?

RQ 2: What are the key drivers of energy consumption in AI-based educational systems?

RQ 3: How much energy can be conserved by choosing the optimal model concerning energy consumption?

We analyze the energy consumption of four alter-

[a] https://orcid.org/0009-0001-6941-1612
[b] https://orcid.org/0000-0001-6792-461X
[c] https://orcid.org/0000-0003-3123-1517
[d] https://orcid.org/0000-0003-1595-175X

Bültemann, M., Simbeck, K., Rzepka, N. and Kalff, Y.
Sustainable Learning Analytics: Measuring and Understanding the Drivers of Energy Consumption of AI in Education.
DOI: 10.5220/0012547600003693

In *Proceedings of the 16th International Conference on Computer Supported Education (CSEDU 2024) - Volume 1*, pages 272-279
ISBN: 978-989-758-697-2; ISSN: 2184-5026

native ML models designed to predict the likelihood of correctly solving a task. Training and evaluating models based primarily on performance metrics like accuracy and speed is standard practice. Energy consumption usually is not a determining factor. In this study, we assess the energy consumption of models for potential integration into a learning platform serving several thousand users. All models were trained using the same dataset. The energy consumption of the models was measured using the PowerJoular library (Noureddine, 2022).

This study specifically targets the energy consumption associated with model inference. The energy needs for training the model, and the learning platform itself are ignored — so are life-cycle emissions, including hardware production, use and disposal, and resources from software production and uninstallation (Kern et al., 2018).

2 LITERATURE REVIEW

2.1 Measuring Energy Consumption of Software

The energy consumption of software can be gauged through both hardware and software-based methodologies. Noureddine et al. (2012) propose a software-based approach and introduce the PowerAPI library, which measures the energy consumption on the operating system level. Raw information is collected from sensor modules (e.g., central processing unit (CPU) or network) and used to calculate the system power consumption as well as the power consumption of single running processes. The authors test their approach by running different algorithm implementations and conclude that an algorithm's choice affects an application's energy consumption. A software-based tool introduced more recently is PowerJoular (Noureddine, 2022), which uses interfaces provided by hardware manufacturers.

An example of a hardware-based approach is to use a measuring station consisting of two PCs - one to execute the system under test and one to collect the measured energy data (Junger et al., 2022). The method is applied by Bültemann et al. (2023) to measure the energy usage of different AIED (AI in education) algorithms. Verdecchia et al. (2018) implement a similar approach and extend it to a spectrum-based energy hotspot localization to identify energy-intensive components in the code. Mancebo et al. (2021) propose a general process that consists of a hardware component that measures the energy consumption and a software application that analyzes the results. This proposal is applied in a different study on the energy utilization of health data (Garcia-Berna et al., 2021).

Kern et al. (2018) suggest assessment criteria when measuring resource and energy consumption of software usage. These include user autonomy to configure resource-saving settings, resource management and default settings, hardware requirements (e.g., electricity consumption), and backward compatibility when releasing new versions. While numerous studies have been conducted in the past years, and many processes share similarities, there is currently no universally recognized standard process for energy assessments of software.

2.2 Measuring Energy Consumption of Machine Learning

Machine learning is used for various cases like speech and image recognition, translations, or recommendations. ML algorithms use large data sets to recognize patterns, learn a behavior, and make predictions (Helm et al., 2020). Several parameters contribute to the total energy costs of AI: execution costs of a model, size of the training data set, and the total number of hyperparameter experiments (i.e., number of training iterations) (Schwartz et al., 2020; Sharir et al., 2020). The increasing size of a model directly increases the costs for training and inference. In addition, energy utilization of storage is rising as the data sets become increasingly larger (Dhar, 2020). Other drivers of the environmental cost of ML are the local energy mix, water demand for data center cooling, and electronic waste (Patterson et al., 2021; Wu et al., 2022; Henderson et al., 2020). To reduce the environmental impact of ML models, the choice of models and hardware is essential (Schwartz et al., 2020; Sharir et al., 2020).

2.3 Artificial Intelligence and Learning Analytics

LA is "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs" (LAK, 2011; Greller and Drachsler, 2012). Patterns in a learner's behavior are recognized, and individual measures may be drawn (Gobert and Sao Pedro, 2016). LA is implemented in massive open online courses, intelligent tutoring systems, or learning management systems (Lu et al., 2018).

Machine learning in learning analytics has grown considerably in the last decade. AI for Learning

Analytics relies on large amounts of data to generate models and predictions (Brooks and Thompson, 2017). Data is collected from the interaction between instructors, learners, the educational environment, and administrative or performance processes (Brooks and Thompson, 2017; Romero et al., 2014). Depending on the use case, AI techniques such as machine learning, natural language processing, or recommender systems are applied in LA (Brooks and Thompson, 2017; McNamara et al., 2017; Chen et al., 2020). Chen et al. (2020) investigate the impact of AI on the education sector by reviewing 30 studies. They find that AI in education can take different forms, e.g., in computers, embedded systems, or web-based platforms. It primarily supports instructors pursuing administrative tasks like grading or providing feedback. On the other hand, students benefit from AIED as the educational content is tailored to their individual needs, thus increasing learning success. They conclude that AI has a considerable effect on the educational sector. The use of AI in LA enables adaptivity and individualized interventions. An adaptive learning environment observes and interprets the user's actions to identify and model learners' individual needs and preferences (Paramythis and Loidl-Reisinger, 2003). AI adapts to a user's unique level of skills in real time and supports the personal learning process using educational data. Interventions take place in the forms of direct messages to the learner (i.e., when a learner is identified as being at risk of failing), individual content, or actionable feedback (Wong and Li, 2018).

3 CASE STUDY: THE LEARNING PLATFORM "Orthografietrainer"

Orthografietrainer[1] (OT) is a learning platform for students from grades 5 to 10 to acquire and consolidate German spelling skills. It was established in 2008 and has since been accessed by more than 1.25 million. users. The platform provides more than 5000 tasks, which have been solved over 12 million times.

The training platform offers exercises in various areas of competencies, such as comma formation or capitalization. Each set of exercises bundles ten sentences that have to be completed. In its default setup, the platform works as follows: if a learner enters a sentence incorrectly, two additional sentences will be added to the exercise set before the previously wrongly submitted sentence is displayed again. It provides immediate feedback and dynamic adjustments. In the scope of an experiment, OT was transformed into an adaptive platform using a machine learning model to predict the probability of solving a task sentence correctly (Rzepka et al., 2022, 2023). The deployed adaptive learning interventions on the platform outperform the original setup regarding learning gains and drop-outs (Rzepka, 2023). Before conducting the experiment, four different ML algorithms were trained on the data set: a decision tree, a logistic regression, SVM, and MLP. The dependent variable is the success of solving a task correctly (1) or incorrectly (0). The data set consists of (1) demographic data (e.g., grade and gender), (2) task data (e.g., difficulty, previous attempts), and (3) a learning history of submitted exercises over the past three months (Rzepka, 2023). The models differ slightly in accuracy (96.62 - 97.09), precision (97.38 - 97.97), and recall (95.82 - 96.31) (Rzepka, 2023).

[1] https://orthografietrainer.net/

4 EXPERIMENTAL SETUP

4.1 Experimental Setup

In this study, we measure the energy consumption of model inference, i.e., when the trained model is provided with new data to predict the probability of solving a sentence correctly. In each experiment, the Python code predicts the probability of doing the task using one of the models. Each model is executed 30 times, and an average is calculated for the measured values. The averages of the four models are compared to the baseline (BL), which is the metered energy consumption of the system without inference running. Before the experiments, the computer was newly set up with the operating system, required software, and dependencies for the measurement. Because of the model requirements, Python is installed in two different versions: version 3.7 for the decision tree, logistic regression, and SVM, and version 3.10 for the MLP. Only local resources are measured in this setup, as it is impossible to access any further infrastructure. The model is saved locally, so the differences across models can be estimated. The PowerJoular software (Noureddine, 2022) measures energy and system resource consumption. It is deployed on a Raspberry Pi. PowerJoular reads resource consumption on the CPU cycles using a power polynomial regression model read from the `/proc/stat` system file. Its error rate is estimated between 0.3% and 3.83%. PowerJoular can monitor the entire system or processes based on a process ID.

Measurements are conducted on a Raspberry Pi 4b with 4 GB RAM with Broadcom BCM2711, a 64-bit system on chip, and a processor sub-architecture at 1.5GHz. Raspberry Pi devices have been used for various studies in the past (Kumar et al., 2018; Saha et al., 2018; Umarghanis et al., 2020) as they are inexpensive and compatible with many other hardware devices.

In this setup, PowerJoular is used to measure the system's CPU usage and electric power. Memory usage is captured by the Linux command line tool PS, as suggested by Wagner et al. (2023). PS measures the relative memory usage of the system on a process basis (Balister et al., 2007). In this experiment, the processes of each model execution (i.e., the execution of the respective Python script) are monitored, and memory is reported. In the baseline measurement, the entire system is monitored. Time stamps are logged in the code to get an accurate start and stop time for each of the 30 measurement cycles. The variables gathered are CPU usage as a percentage, memory consumption as a percentage, and electrical power consumption in watts.

4.2 Experimental Hypotheses

The objective of the following measurements is to analyze the effect of the independent variables (the models, i.e., decision tree, logistic regression, SVM, and MLP) on the dependent variables (CPU usage, memory usage, power usage) to answer RQ1 and RQ2. The null hypothesis states that there is no significant difference between the independent variables (m) and the dependent variables. The alternative hypothesis H_1 states that a statistically significant difference between a dependent and an independent variable is observed.

$$H_0 : \mu_m = \mu_{\text{BL}}$$
$$H_1 : \mu_m \neq \mu_{\text{BL}}$$

RQ3 aims to determine how much energy could be conserved by choosing a model. We assume that the amount of energy saved in comparison to the baseline is not significant (H_0), as opposed to statistically significant savings (H_2).

$$H_0 : \mu_m = \mu_{\text{BL}}$$
$$H_2 : \mu_m \neq \mu_{\text{BL}}$$

5 DATA AND RESULTS

5.1 Measured System Demand of ML Models and Baseline System

We measure system demand using CPU utilization [%], memory utilization [%], and power in Watt (W)[2]. The values of these measures for all four models and the baseline system are available in table 1. Compared to the baseline average of 1.8%, CPU utilization measured by PowerJoular of the four models ranges between 24.47% for the decision tree (DT) model and 33.59% for the multilayer perceptron (MLP) model. The lowest mean value for memory utilization is also measured in the decision tree with 29.47%, the highest in MLP (41.97%). Baseline memory utilization is 7%. For electric power, mean values range from 3.33W/3.38W (decision tree/logistic regression) to 3.91W (MLP), in contrast to a baseline metric of 3.24W.

Table 1: Descriptive variables for system demand.

	Min	Max	Mean	Median	Std
CPU Utilization DT [%]	1.00	33.84	24.47	26.75	0.07
Memory Utilization DT [%]	1.20	41.8	29.74	31.8	8.12
Electric Power DT [W]	3.03	3.65	3.33	3.37	0.1
CPU Utilization LogReg [%]	1.5	52.12	26.1	26.82	0.06
Memory Utilization LogReg [%]	15.5	41.9	32.72	33.75	7.19
Electric Power LogReg [W]	3.03	4.70	3.38	3.36	0.19
CPU Utilization SVM [%]	1.25	79.34	28.36	28.25	0.02
Memory Utilization SVM [%]	1.20	43.9	30.92	30.9	1.13
Electric Power SVM [W]	3.03	5.86	3.42	3.41	0.08
CPU Utilization MLP [%]	0.0	50.50	33.59	45.89	0.17
Memory Utilization MLP [%]	8.5	63.0	41.97	43.05	10.01
Electric Power MLP [W]	3.03	4.61	3.91	4.33	0.6
Memory Utilization BL [%]	7.0	8.0	7.00	7.0	0.01
Electric Power BL [W]	3.03	3.41	3.24	3.26	0.05

In the next step, the mean values are compared to the baseline values (table 2). All models require more resources in terms of CPU utilization, memory utilization, and electric power than the baseline.

[2]Cf. the GitHub repository https://github.com/marlenebuelt/messung-2 for results.

Table 2: Relative difference to baseline values (pp.: percentage points).

	Delta DT %	Delta LogReg %	Delta SVM %	Delta MLP %
CPU Utilization [pp.]	1260.49	1350.90	1476.93	1767.92
Memory Utilization [pp.]	324.90	367.49	341.72	499.53
Electric Power [pp.]	2.90	4.44	5.65	20.63

The data undergoes a Shapiro-Wilk test to assess normality. However, the results indicate that the data is not normally distributed. Therefore, a Mann-Whitney U test compares the mean system demand parameters between the models and baseline. The effect sizes are estimated using Cliff's delta. It takes on a value between -1 and 1, with 0 indicating no effect and $|1|$ a very strong effect. The threshold for a medium effect is set at a value larger than $|0.28|$, and a large effect is considered $|0.43|$ and above (Vargha and Delaney, 2000). Table 3 shows that almost all differences between models and between baseline and models are highly significant with a p-value of 0.00. Only the p-values for the correlation of CPU utilization and power conversion between DT and LogReg models are not close to zero but are still very low at 0.01. As indicated by Cliff's delta (refer to Table 4), the effect size consistently demonstrates a high magnitude when comparing the ML models to the baseline. Additionally, there is a notable high effect size between the decision tree and logistic regression models on the one hand and SVM and MLP on the other.

Table 3: Mann–Whitney U test.

	BL	DT	LogReg	SVM
CPU				
DT	**0.00**			
LogReg	**0.00**	**0.01**		
SVM	**0.00**	**0.00**	**0.00**	
MLP	**0.00**	**0.00**	**0.00**	**0.00**
Memory				
DT	**0.00**			
LogReg	**0.00**	**0.00**		
SVM	**0.00**	**0.00**	**0.00**	
MLP	**0.00**	**0.00**	**0.00**	**0.00**
El. Power				
DT	**0.00**			
LogReg	**0.00**	**0.01**		
SVM	**0.00**	**0.00**	**0.00**	
MLP	**0.00**	**0.00**	**0.00**	**0.00**

Table 4: Cliff's delta.

	BL	DT	LogReg	SVM
CPU				
DT	**-0.98**			
LogReg	**-0.98**	-0.11		
SVM	**-1.00**	**-0.76**	**-0.70**	
MLP	**-0.91**	**-0.41**	**-0.37**	-0.21
Memory				
DT	**-0.97**			
LogReg	**-1.00**	-0.23		
SVM	**-1.00**	0.12	**-0.7**	
MLP	**-1.00**	**-0.75**	**-0.37**	**-0.81**
El. Power				
DT	**-0.69**			
LogReg	**-0.85**	-0.11		
SVM	**-0.99**	**-0.76**	**-0.70**	
MLP	**-0.62**	**-0.44**	**-0.38**	-0.21

5.2 Estimation of the Annual Energy Consumption

To interpret the system demand, the model runtimes need to be considered. The models differed in mean power and average run time per inference (table 5). To calculate the energy consumption per year, we require an estimation of the total number of inferences per year. We base our assumption on the experiment described in Rzepka (2023). In four months, 38,000 sessions took place on the platform (114,000 sessions per year). One session is assumed to include an exercise set of ten sentences. As the first sentence is predefined, nine inferences are undertaken to determine the second sentence, eight inferences for the third, and so forth. This sums up to 44 inferences. However, if the prediction for a learner to correctly solve a sentence is determined to be less than 50%, the procedure changes, and additional inferences are executed to provide training sentences. Hence, the total number of inferences in a session varies. This paper assumes 64 inferences per session, thus 7,296,000 inferences per year.

Table 5 summarizes the use of electric power in a single inference in watt-seconds (Ws), the estimated usage of power per year in watt hours (Wh) as explained in chapter 5.2, and the costs of electric power in euros per year. We assume the average market price for electricity in Germany for the year 2022 of 0.3279 EUR per kWh Eurostat (2023). The average usage per inference ranges between 45.77 Ws in the decision tree model and 3,783.27 Ws when executing an SVM. This sums up to 92,760 Wh in a decision tree

Table 5: Estimation of annual energy consumption.

	Time per Inference [s]	Mean Power [W]	Energy per Inference [Ws]
DT	13.8	3.33	45.77
LogReg	15.97	3.38	51.64
SVM	4085.00	3.42	3783.27
MLP	34.83	3.91	129.19
	Energy per Year [kWh]	**Costs per Year [€]**	
DT	92.8	30.43	
LogReg	104.7	34.33	
SVM	7667.4	2514.14	
MLP	261.8	85.84	

and to 7,667,428 Wh in an SVM. The energy costs of a decision tree amount to 30.43 Eur, of a logistic regression 34.33 Eur, of an SVM 2,514.14 Eur, and of an MLP 85,84 Eur.

6 DISCUSSION

Answering RQ1, we find that ML models vary in energy consumption. The data indicates significant differences between all models and the baseline as well as between the models. The low basic load of a Raspberry Pi contributes to the significance and the effect sizes when compared against the baseline. Our data indicates that the decision tree is slightly more efficient than the logistic regression model, while MLP and especially SVM require much more electric power. The similar significance values and effect sizes of CPU utilization and electric power consumption in the PowerJoular measurement indicate that the parameters correlate. Thus, CPU usage could be considered as a proxy for power use.

The general findings of this experiment coincide with other findings, e.g., of Strubell et al. (2019) and Frey et al. (2022). The study reveals that differences between models are discernible and exhibit a broad range. These findings align with the authors' conclusions drawn in a prior experiment, as documented in (Bültemann et al., 2023).

Our results demonstrate that the complexity of a model contributes to the consumption of computational resources and, thus, energy consumption (RQ2). The execution of all models resulted in a significant increase in the load on the observed parameters (CPU utilization, memory utilization, electric power) compared to the baseline. This effect was most pronounced for SVM and MLP. Less complex models, such as decision trees and logistic regressions, could make predictions more rapidly than SVM and MLP. Several layers and the number of interconnected neurons contribute to the increased energy consumption of an MLP. SVMs require more runtime and energy due to the complex optimization process of finding the largest margin between many data points. The training kernel matrix squares with the size of the dataset, thus making it inefficient on large data sets like the ones used in this experiment. On the other hand, decision trees and logistic regressions have a relatively simple prediction process that is boolean or mathematically explainable. Memory usage reveals the highest values in MLP. This indicates that a large amount of data is processed in the prediction process of the models. Regarding RQ3, our results indicate that the SVM model would substantially increase energy consumption by 30-80. Real-life implications could differ based on the hardware used in the actual data center, but we expect an effect of the same magnitude.

It is essential to acknowledge certain limitations that reduce the generalizability of our findings. This study only analyses ML model inference's energy demand and neglects model training's energy consumption and life-cycle emissions. Further, results depend on the deployed hardware; data centers usually utilize different hardware. Finally, models should also be trained on other datasets to validate results.

As we reveal differences in the resource consumption of different ML models, future research in ML implementations for LA should consider the resource impact of the chosen models in line with parameters such as accuracy, fairness, and explainability. Both energy consumption and learning outcomes should be considered when comparing AI-related emissions and learning outcomes to find a well-balanced, cost-efficient solution. Energy costs, in particular, are an essential economic determinant in times of persistently high energy prices. If the performance of models reveals similar results, environmental aspects should be considered. This confirms the findings in Schwartz et al. (2020), who propose to focus AI research on efficiency measures.

7 CONCLUSION

This study aimed to fill a critical research gap by investigating the environmental impact of machine learning models on educational platforms, specifically focusing on their energy efficiency and consumption. Our findings indicate significant variations in energy consumption across different ML models. Decision trees and logistic regression models are more energy-efficient than SVMs and MLPs. The study also reveals that the model's complexity is proportional to its energy consumption, with SVM and MLP requiring considerably more computational resources and computing time. These insights answer our research questions, providing a better understanding of the critical drivers of energy consumption in AI-based educational systems (RQ2) and the potential energy savings by selecting more efficient models (RQ3).

From a practical perspective, our findings indicate that choosing an energy-efficient model could reduce energy consumption significantly compared to more resource-intensive models. This is particularly relevant for educational platforms that operate at scale, where even marginal improvements in energy efficiency can lead to substantial reductions in operational costs and carbon footprint. A key driver is a model's complexity caused by the training algorithms. A logistic regression is mathematically explainable, while a decision tree uses a boolean decision process. SVMs use complex optimization processes, and MLPs consist of many layers of interconnected neurons, which makes prediction processes more complex and the model larger in storage. This causes the processes' resource consumption to rise and increases the runtime. Our work has shown that choosing resource-efficient ML models for educational purposes is complementary to achieving objectives related to explainability and speed of inference.

Future research could extend this work by considering a broader range of ML models and algorithms. Studies could also incorporate energy consumption during the training phase and other life-cycle emissions to provide a more comprehensive view of the environmental impact. This study shows options for incorporating sustainability considerations into developing LA systems. Performance predictions imply that significantly lower energy consumption can achieve the same results. In conclusion, educational achievements and environmental sustainability are not mutually exclusive.

REFERENCES

(2011). *LAK '11: Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, New York. ACM.

Alharthi, A. D., Spichkova, M., and Hamilton, M. (2019). Sustainability requirements for eLearning systems: a systematic literature review and analysis. *Requirements Engineering*, 24:523–543.

Anser, M. K., Ahmad, M., Khan, M. A., Zaman, K., Nassani, A. A., Askar, S. E., Abro, M. M. Q., and Kabbani, A. (2021). The role of information and communication technologies in mitigating carbon emissions: evidence from panel quantile regression. *Environmental Science and Pollution Research*, 28:21065–21084.

Balister, P. J., Dietrich, C., and Reed, J. H. (2007). Memory usage of a software communication architecture waveform. In *2007 Software Defined Radio Technical Conf. Product Exposition*. Citeseer.

Brooks, C. and Thompson, C. (2017). Predictive modelling in teaching and learning. *Handbook of learning analytics*, pages 61–68.

Bültemann, M., Rzepka, N., Junger, D., Simbeck, K., and Müller, H.-G. (2023). Energy consumption of ai in education: A case study. In *21. Fachtagung Bildungstechnologien (DELFI)*, pages 219–224. Gesellschaft für Informatik e.V., Bonn.

Calero, C. and Piattini, M. (2017). Puzzling out software sustainability. *Sustainable Computing: Informatics and Systems*, 16:117–124.

Chen, L., Chen, P., and Lin, Z. (2020). Artificial Intelligence in Education: A Review. *IEEE Access*, 8:75264–75278.

Dhar, P. (2020). The carbon impact of artificial intelligence. *Nature Machine Intelligence*, 2(8):423–425.

Durdik, Z., Klatt, B., Koziolek, H., Krogmann, K., Stammel, J., and Weiss, R. (2012). Sustainability guidelines for long-living software systems. In *2012 28th IEEE International Conference on Software Maintenance (ICSM)*, pages 517–526.

Eurostat (2023). Electricity prices by type of user. *Data Browser*.

Frey, N. C., Li, B., McDonald, J., Zhao, D., Jones, M., Bestor, D., Tiwari, D., Gadepally, V., and Samsi, S. (2022). Benchmarking resource usage for efficient distributed deep learning. In *2022 IEEE High Performance Extreme Computing Conference (HPEC)*, pages 1–8. IEEE.

Garcia-Berna, J. A., Fernandez-Aleman, J. L., de Gea, J. M. C., Toval, A., Mancebo, J., Calero, C., and Garcia, F. (2021). Energy efficiency in software: A case study on sustainability in personal health records. *Journal of cleaner production*, 282:124262.

Gobert, J. D. and Sao Pedro, M. A. (2016). Digital assessment environments for scientific inquiry practices. *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications*, pages 508–534.

Greller, W. and Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society*, 15(3):42–57.

Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., Spitzer, A. I., and Ramkumar, P. N. (2020). Machine learning and artificial intelligence: definitions, applications, and future directions. *Current reviews in musculoskeletal medicine*, 13:69–76.

Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D., and Pineau, J. (2020). Towards the systematic reporting of the energy and carbon footprints of machine learning. *The Journal of Machine Learning Research*, 21(1):10039–10081.

IPCC (2022). *Summary for Policymakers*, page In Press. Cambridge University Press, Cambridge, UK. Editors: Pörtner, H. O. and Roberts, D. C. and Tignor, M. and Poloczanska, E. S. and Mintenbeck, K. and Alegría, A. and Craig, M. and Langsdorf, S. and Löschke, S. and Möller, V. and Okem, A. and Rama, B.

Junger, D., Wohlgemuth, V., and Kammer, E. (2022). Conception and test of a measuring station for the analysis of the resource and energy consumption of material flow-oriented environmental management information systems (EMIS). In Wohlgemuth, V., Naumann, S., Arndt, H.-K., Behrens, G., and Höb, M., editors, *EnviroInfo 2022*, page 211, Bonn. Gesellschaft für Informatik e.V.

Kern, E., Hilty, L. M., Guldner, A., Maksimov, Y. V., Filler, A., Gröger, J., and Naumann, S. (2018). Sustainable software products—Towards assessment criteria for resource and energy efficiency. *Future Generation Computer Systems*, 86:199–210.

Kumar, A., Chattree, G., and Periyasamy, S. (2018). Smart healthcare monitoring system. *Wireless Personal Communications*, 101:453–463.

Lu, O. H., Huang, A. Y., Huang, J. C., Lin, A. J., Ogata, H., and Yang, S. J. (2018). Applying learning analytics for the early prediction of Students' academic performance in blended learning. *Journal of Educational Technology & Society*, 21(2):220–232.

Mancebo, J., Garcia, F., and Calero, C. (2021). A process for analysing the energy efficiency of software. *Information and Software Technology*, 134:106560.

McNamara, D. S., Allen, L., Crossley, S., Dascalu, M., and Perret, C. A. (2017). Natural language processing and learning analytics. *Handbook of learning analytics*, 93.

Naumann, S., Guldner, A., and Kern, E. (2021). The eco-label blue angel for software—Development and components. In *Advances and New Trends in Environmental Informatics: Digital Twins for Sustainability*, pages 79–89. Springer.

Noureddine, A. (2022). PowerJoular and JoularJX: multi-platform software power monitoring tools. In *2022 18th International Conference on Intelligent Environments (IE)*, pages 1–4. IEEE.

Noureddine, A., Bourdon, A., Rouvoy, R., and Seinturier, L. (2012). A Preliminary Study of the Impact of Software Engineering on GreenIT. In *2012 First International Workshop on Green and Sustainable Software (GREENS)*, pages 21–27. IEEE.

Paramythis, A. and Loidl-Reisinger, S. (2003). Adaptive learning environments and e-learning standards. In *Second european conference on e-learning*, volume 1, pages 369–379.

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., So, D., Texier, M., and Dean, J. (2021). Carbon emissions and large neural network training.

Romero, C., Romero, J. R., and Ventura, S. (2014). A survey on pre-processing educational data. *Educational data mining: applications and trends*, pages 29–64.

Rzepka, N. (2023). *Transforming first language learning platforms towards adaptivity and fairness: models, interventions and architecture*. PhD thesis, Humboldt Universiät zu Berlin, Berlin.

Rzepka, N., Simbeck, K., Müller, H.-G., and Pinkwart, N. (2022). Keep it up: In-session dropout prediction to support blended classroom scenarios. In *CSEDU (2)*, pages 131–138.

Rzepka, N., Simbeck, K., Müller, H.-G., and Pinkwart, N. (2023). Go with the Flow: Personalized Task Sequencing Improves Online Language Learning. In *International Conference on Artificial Intelligence in Education*, pages 90–101. Springer.

Saha, A. K., Sircar, S., Chatterjee, P., Dutta, S., Mitra, A., Chatterjee, A., Chattopadhyay, S. P., and Saha, H. N. (2018). A raspberry Pi controlled cloud based air and sound pollution monitoring system with temperature and humidity sensing. In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 607–611. IEEE.

Schwartz, R., Dodge, J., Smith, N. A., and Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12):54–63.

Sharir, O., Peleg, B., and Shoham, Y. (2020). The cost of training nlp models: A concise overview. *arXiv preprint arXiv:2004.08900*.

Stroustrup, B. (2014). *Programming: principles and practice using C++*. Pearson Education.

Strubell, E., Ganesh, A., and McCallum, A. (2019). Energy and policy considerations for deep learning in NLP.

Umarghanis, S. A., Darari, F., and Wibisono, A. (2020). A Low-Cost IoT Platform for Crowd Density Detection in Jakarta Commuter Line. In *2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pages 121–128. IEEE.

Vargha, A. and Delaney, H. D. (2000). A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics*, 25(2):101–132.

Verdecchia, R., Guldner, A., Becker, Y., and Kern, E. (2018). Code-level energy hotspot localization via naive spectrum based testing. In *Advances and New Trends in Environmental Informatics: Managing Disruption, Big Data and Open Science*, pages 111–130. Springer.

Wagner, L., Mayer, M., Marino, A., Nezhad, A. S., Zwaan, H., and Malavolta, I. (2023). On the Energy Consumption and Performance of WebAssembly Binaries across Programming Languages and Runtimes in IoT. In *Proceedings of the 9th International Conference on Evaluation and Assessment on Software Engineering (EASE)*.

Wong, B. T. and Li, K. C. (2018). Learning analytics intervention: A review of case studies. In *2018 International Symposium on Educational Technology (ISET)*, pages 178–182. IEEE.

Wu, C.-J., Raghavendra, R., Gupta, U., Acun, B., Ardalani, N., Maeng, K., Chang, G., Aga, F., Huang, J., Bai, C., et al. (2022). Sustainable ai: Environmental implications, challenges and opportunities. *Proceedings of Machine Learning and Systems*, 4:795–813.