Large Scale Machine Learning Applications in Industries: An Exploration of Energy Consumption

ZIAD ELLEITHY, University of Twente, The Netherlands

Large scale machine learning is on the rise due to the various advancements in technology. Multiple industries are pushing to become smarter as fast as possible. This paper summarizes the use of large scale machine learning in industries as well as the factors affecting energy consumption of machine learning models. Furthermore, simulations are performed to estimate the amount of energy consumption of training a model. This is extended to include different types of algorithms used in specific large-scale machine learning applications.

Additional Key Words and Phrases: Large Scale Machine Learning, Industry 4.0, Green data mining, energy efficiency

1 INTRODUCTION

The manufacturing industries have had multiple revolutions leading up to now, where the fourth industrial revolution (also known as Industry 4.0 or Smart Industry) is on the rise [15]. This revolution brings forth the transformation of organizations and companies to a digital embodiment which alters how they operate and function [30]. This is mainly achieved through the application of the Internet of Things (IoT) and Cyber-physical Systems (CPS) to link the real and digital structures which allows said organizations to improve their performance if utilized correctly [23].

Through the use of sensors in industries, IoT is enhanced by gathering more and more data. Data can then be processed to provide useful information by Machine Learning (ML) models, which is a sub division of Artificial Intelligence (AI). One of ML models' current weaknesses is the huge time costs when used on large scale data [38]. An estimate was made that around 1 trillion sensors will be used by humanity in 2025 [25]. This will contribute heavily to Big Data, which leads to the necessity of large-scale Machine Learning .

Energy consumption has not been considered much in machine learning research but rather accuracy of the models was the main and most important factor [7]. There is also a growing pressure to be more sustainable especially with the international treaty on climate change that was signed back in 2015 [34]. In the follow up meeting of the treaty in 2021, it was estimated that , with extra implementations to reduce greenhouse gas emissions, the total emissions of the involved parties would be 11.3 percent lower than their goal in 2030 [35]. This provides further reason for machine learning research to take a look at energy consumption as an important factor which [12] did. The energy usage of specific machine learning models was measured and compared with other models while also using two different types of datasets to compare even further.

This leads to the research question:

How much energy would large scale machine learning consume in industry applications?

With the help of these sub-questions this research question can be answered:

(1) What is massive scale machine learning in industries and what are its applications?
(2) What are the factors that affect energy consumption of machine learning processes?

This paper aims to be a hub for data scientists in the industry aiming to be more aware of the different large scale machine learning processes as well as the factors that contribute to their energy consumption.

2 METHODOLOGY

This research will be conducted mainly of qualitative analysis by means of a systematic literature review that loosely follows the guidelines of Kitchenham [14] and Wohlin [40]. This method of literature review helps identify, evaluate and interpret the state of the art research relevant to the research questions to be answered in this paper [14]. In Wohlin's paper, it is suggested to identify a set of papers that start as a starting point which are then used to snowball [40]. Therefore, known item search will be carried out first to find the most relevant papers which would then be used to snowball to reduce the number of irrelevant studies by using papers that have been cited in said relevant papers.

All database searches are done through Scopus and Web of Science for sub question 1 as they contains a wide variety of articles and journals. For the first subquestion, the search terms used are "TITLE-ABS-KEY (("large scale machine learning" OR "massive scale machine learning" OR "large-scale machine learning") AND ("industry 4.0" OR "smart industr*" OR "industr*"))" as well as "("large scale machine learning" OR "massive scale machine learning" OR "large-scale machine learning") AND ("industry 4.0" OR "smart industr*" OR "industr*")" for Scopus and Web of Science respectively. Take a look at Table 1 to have a better overview of the search results. This search came up with 24 and 26 results respectively for Scopus and Web of Science. Before reading any of them all duplicates were removed which amounted to a total of 15 duplicates. Irrelevant papers were also removed if they talked only about the optimization of specific algorithms or did not mention which machine learning models were used in an application. From these results, one key document called "Large-scale machine learning systems in real-world industrial settings: A review of challenges and solutions" [20] was used to further snowball and find relevant applications in industries. This snowball resulted in 18 additional relevant papers.

For sub question 2, it was difficult to conduct a systematic literature review as not a lot of papers can be easily found regarding through Scopus and Web of Science with the search terms "Green

TScIT 37, July 8, 2022, Enschede, The Netherlands

Table 1. Search Terms Results

Question	Search Term	Results	Duplicates	Irrelevant	Snowball
Sub Question 1	("large scale machine learning" OR "massive scale machine learning" OR "large-scale machine learning") AND ("industry 4.0" OR "smart industr*" OR "industr*"))	50	15	23	18

AI" or ("factor*" AND "energy consumption" and ("machine learning" or "AI")) . Only one paper was found through Scopus, that contributes to the question, called "Green AI" [29]. Therefore a non systematic literature review was conducted that resulted in a total of 5 papers that discuss energy consumption in machine learning. Through the help of the supervisor [12, 28] were provided at the start of the thesis and [2] was found through Google Scholar. The final paper was found by searching through the Vrije Universiteit Amsterdam's Database of the Software and Sustainability research group [36] which is what this paper builds upon for the model and simulations .

Furthermore, a small quantitative analysis will be established to help answer the main research question by estimating roughly, through other literature work, how much energy would be consumed by specific applications. To achieve this modelling and simulations will be used for which Insight Maker will be utilized. Insight Maker is an open-source modelling web-based tool that focuses on accessibility and the inclusion of various features [10].

3 RESULTS

In this section each sub question will first be answered individually after which the main question will be asnwered.

3.1 Sub Question 1

The structure behind machine learning implementations is that their logic is generated by continuously learning from data instead of the logic being explicitly programmed in traditional software [20]. Moreover what differs large scale machine learning from regular machine learning is the large amount of data, having billions of instances features or classes [5]. In the following paragraphs, the different applications of large scale machine learning will be listed along with the used/preferred machine learning model (see Table 2 for an overview).

Recommendation systems are a popular application of large scale machine learning used in online applications such as Alibaba, the Facebook Marketplace, and Youtube [6, 8, 11, 37, 45]. The most common machine learning model used in this application is Neural Networks along with either Graph Embedding [8, 37, 45] or Logistic Regression [6]. There is one study that stands out in this field which uses Topic Modelling and Random Forest models to generate user profiles based on what applications they have installed on their phone [11]. The study concludes that the machine learning models used can notably increase correct prediction of a user's interest and gender.

In the advertisement industry, large scale machine learning is utilized for the prediction of click through rates on advertisements or for displaying the right type of advertisements for each user [1, 3, 26]. Linear and Logistic Regression models are used to predict click through rates by using hundreds of terabytes of data to maximize revenue of advertising companies [1, 3]. Perlich et al. also describes how Linear and Logistic Regression models can be used for targeted display advertising. [26]

Fraud, anomaly and intrusion detection systems are another important application of large scale machine learning [17, 31, 39, 43, 46, 47]. For fraud detection, multiple models are viable such as Random Forest, Logistic Regression and Decision Tree models [39, 43, 46]. However, in the case of fraud detection in mobile device payment, Gradient Boosting Decision Trees were proven to be most accurate [46]. In terms of anomaly detection, Support Vector Machine models are used more frequently to analyse sensor data [47]. With regards to online intrusion detection, Support Vector Machines, Logistic Regression, and Random Forest models have been utilized to detect intruders in Facebook and Skype accounts [17, 31]. In Skype's case, it was reported that the Random Forest model performed 10 percent better than other models [17].

Text Mining is another use case of large scale machine learning where all studies used Topic Modelling to categorize text in social media applications such as Twitter [4, 41, 42]

In the telecommunication field two applications were identified through the literature review. The first being dynamic real-time network monitoring using Neural Networks which ensures high reliability and availability of telecommunication networks [22]. The other application is identifying influential subscribers through social network analysis by using Neural Networks, Logistic Regression and Decision Tree models [21].

Random Forest models are employed within the medical industry to help predict which advanced visualization algorithm should be applied in specific scenarios to save time [9].

In an industrial production setting for metal casting, machine learning was used for the creation of quality prediction models [16]. In this process multiple models were used such as Neural Networks, Random Forest, and Support Vector Machines.

Visible Light Positioning, which is a term for all executions of optical wireless indoor positioning systems, can use different types of machine learning models for location based services such as Neural Networks, K-Nearest Neighbours, and Support Vector Machines [27].

Traffic Sign Recognition systems in vehicles use Density-Based Clustering models to construct auto-pilot maps that are used by AI in autonomous vehicles for auto-piloting features [44]

The rubber industry uses Neural Networks to predict vulcanization data of rubber gum for the production of tires[19].

Table 2. ML Application Mapped to ML Models used

		ML Model								
		NN	SVM	KNN	LR	TM	DT	RF	GE	DBC
Application	Warehousing (Location Based Services) [27]	Y	Y	Y	-	-	-	-	-	-
	Text Mining [4, 41, 42]	-	-	-	-	Y	-	-	-	-
	Traffic Sign Recognition [44]	-	-	-	-	-	-	-	-	Y
	Rubber Manufacturing [19]	Y	-	-	-	-	-	-	-	-
	Near-Wall Modeling for Turbulence [32]	Y	-	-	-	-	-	-	-	-
	Job Title and Query Classification [13, 18]	-	Y	Y	-	-	-	-	-	-
	Flight Arrival Time Prediction [24]	-	-	-	Y	-	Y	-	-	-
	Autonomous Pallet Trucks [33]	Y	-	-	-	-	-	-	-	-
	Recommendation Systems [6, 8, 11, 37, 45]	Y	-	-	Y	Y	-	Y	Y	-
	Click through rate & Ad targeting [1, 3, 26]	-	-	-	Y	-	-	-	-	-
	Fraud/Anomaly/Intrusion Detection [17, 31, 39, 43, 46, 47]	Y	Y	-	Y	-	Y	Y	-	-
	Telecommunication [21, 22]	Y	-	-	Y	-	-	Y	-	-
	Advanced Medical Imaging & Scanning [9]	-	-	-	-	-	-	Y	-	-
	Quality Prediction and Operation Control [16]	Y	Y	-	-	-	-	Y	Y	-

Table 3. Table 2 Legend

ML Model	Full Name
NN	Neural Network
SVM	Support Vector Machine
KNN	K-Nearest Neighbours
LR	Linear/Logistic Regression
K-M	K Means Clustering
TM	Topic Model
RF	Random Forest
GE	Graph Embedding
DT	Decision Tree
DBC	Density Based Clustering

Neural Networks are also used for wall functions which are used for near wall modelling for turbulence of air crafts [32].

Another aviation industry usage of machine learning is regarding real-time flight arrival time predictions [24]. For this, regression models as well as Decision Trees are utilized.

Classification of different subjects is another job for machine learning. Support Vector Machines and K-Nearest Neighbours models are used in the HR industry to classify and categorize job titles [13]. While only Support Vector Machines models are used in vehicles to automate the classification of queries from drivers using unconstrained natural language [18].

Last but not least, autonomous pallet trucks make use of Neural Networks to navigate the work floor for pallet movement operations [33].

3.2 Sub Question 2

There are multiple ways to measure efficiency of machine learning, one of them being electricity usage which is the main focus of this research question [29]. An important factor which is identified here is hardware, as electricity consumption of the same models on the different hardware specifications can differ [28, 29]. This is reinforced by the papers that talk about energy consumption in AI so far, by having hardware as a controlled variable in their experiments [2, 12, 36].

While having hardware as a constant factor, one study explores 3 different factors that could affect energy consumption of model training [36]. The factors explored are algorithm used, number of data points and number of data features. The algorithms experimented on were Support Vector Machines, Decision Trees, K-Nearest Neighbours, Random Forest, Adaptive Boost, and Bagging Classifier. Based on the study's results, it was apparent that the algorithms are a factor as different algorithms consume different amounts of energy with the biggest difference being the K-Nearest Neighbours algorithm consuming 99.49% less energy than the Random Forest algorithm. Keeping in mind that the p-values are minimal, the number of data points positively correlate with the amount of energy consumed within every algorithm where the Support Vector Machines algorithm being the strongest correlated with a coefficient of 0.95 and K-Nearest Neighbours is the weakest but still strongly correlated with a coefficient of 0.80. Those results were further analyzed and it was found that reducing data points can lead to a reduction of energy consumption, up to a minimum of 61.72% for K-Nearest Neighbours and a maximum of 92.16% for Random Forest. Finally, the effect of number of features on energy consumption was investigated. The results show that the number of features are either moderately or strongly correlated to energy consumed for all algorithms except for K-Nearest Neighbours with a correlation coefficient of 0.04. However, the coefficient being so low could be highly up to chance due to the p-value being quite high (0.54).

Further studies confirm that algorithms affect the amount of energy consumed [12]. Two different Decision Tree algorithms were compared on two different data sets. The results show that the Very Fast Decision Tree (VFDT) algorithm consumes 200% less energy compared to the Hoeffding Adaptive Tree (HAT) algorithm on a RandomTree data set with only a cost of 0.35% accuracy. However,

on a RandomRBF data set, VFDT consumes 11.8% less energy at the cost of 16% lower accuracy.

Another study explores the effects of hyperparameters on energy consumption on different data sets only using a multilayer perceptron classifier which is a form of Neural Networks [2]. The hyper parameters explored were hidden layer size, activation function, solver, alpha, and max iterations. The study could not deduce any clear patterns except for hidden layer size having a positive correlation with energy consumption and "tanh" (an activation function parameter) is often consuming the most energy.

Table 4. Factors affecting energy consumption and their maximum energy reduction

	Source					
Factors	[2]	[12]	[28]	[29]	[36]	Reduction
Hardware	Y	Y	Y	Y	Y	N/A
Algorithm	-	Y	-	-	Y	99.49%
Number of Datapoints	-	-	-	-	Y	92.16%
Number of Features	-	-	-	-	Y	75.8%
Hyperparameters	Y	-	-	-	-	N/A

A list of factors affecting energy consumption in machine learning (also shown in Table 4) consists of the following, based on the previous studies:

(1) Hardware
(2) Algorithm
(3) Number of data points
(4) Number of features
(5) Hyperparameters

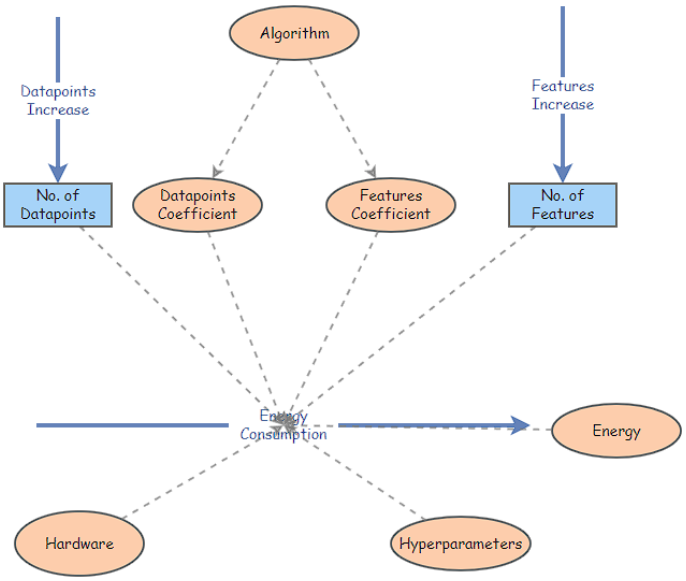

Fig. 1. Model of factors affecting energy consumption of machine learning

Based on these factors a model was created with Insight Maker to visualize said factors which can be see in Figure 1.

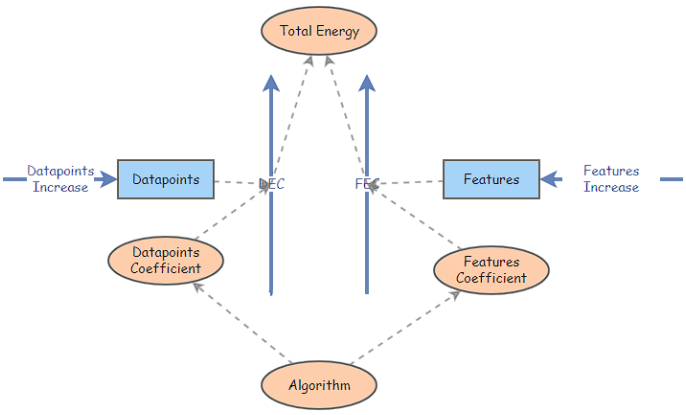

Fig. 2. Adapted Model for Simulation

Table 5. Calculated Model Parameters [36]

Algorithm	Datapoints Coefficient	Features Coefficient
SVM	$1.24 * 10^{-4}$	$9.51 * 10^{-5}$
KNN	$3.75 * 10^{-6}$	$2.06 * 10^{-6}$
DT	$3.25 * 10^{-5}$	$2.19 * 10^{-5}$
RF	$5.01 * 10^{-4}$	$2.96 * 10^{-4}$

3.3 Main Research Question

The model is adapted (Figure 2) into a smaller version as not all factors (hardware and hyperparameters) have been investigated enough, in current literature, to have a clear correlation with energy consumption that can be simulated. The "Energy Consumption" flow (blue arrow) has been divided into 2 different flows called "Datapoints Energy Consumption" and "Features Energy Consumption" For compactness' sake, they were renamed to "DEC" and "FEC" respectively for the model.

For the simulations, the coefficient for datapoints and features are calculated with the help of data gathered from the experiments of Verdecchia et al. [36]. The data was extracted from their github posted in the paper. One thing to keep in mind is that these experiments were run on a constant hardware setup that is a 2.4GHz Quad-Core i5 processor with 16 GB 2133 MHz LPDDR3 RAM. The coefficients were calculated by taking the average amount of energy it took to train the algorithm per set amount of data points or features, then calculating the slope of the best fitting line using Microsoft Excel's LINEST function. The results are in Table 5.

Based on the literature reviewed in sub question 1, only 5 papers included the number of datapoints used as well as the number of features [6, 11, 13, 43, 46]. However one of them uses algorithms that do not have sufficient data to simulate at the moment [6]. .The simulations will extrapolate how much energy would be consumed in large scale machine learning cases that have explicitly state how much datapoints and features they used,while also only considering the four different algorithms that are SVM, KNN, RF, and DT. In the following paragraphs each papers' parameters (number of

Table 6. Estimated Energy Consumption of Machine Learning Applications

Algorithm	Application	Number of Datapoints	Number of Features	Total Energy	Source
SVM	Job Title Classification	2,000,000	280	248.03 J	[13]
KNN	Job Title Classification	2,000,000	280	7.50 J	[13]
DT	Bankcard Enrollment Fraud Detection	100,000	614	3.26 J	[46]
RF	Bankcard Enrollment Fraud Detection	100,000	614	50.28 J	[46]
RF	Mobile Recommendation System	904	1000	0.75 J	[11]
RF	Automatic Cash-Out Fraud Detection	131,000,000	300	65631.09 J	[43]

datapoints and number of features) will be simulated and displayed. The results of the simulations can be seen in Table 6.

For SVM and KNN both simulations (see Table 7 & 8) are relating to Job Title Classification with a training data set of 2 million datapoints and 280 non zero features [13]. With a SVM algorithm, 2 million datapoints would consume around 248 joules and 280 features would add around 0.0266 joules. Meanwhile a KNN algorithm would consume only 7.50 joules for 2 million datapoints and about 0.000618 joules for 280 features.

For the case of Fraud detection within bankcard enrollment, only 100 thousand training data points with 614 features utilizing both an RF and DT algorithm [46]. For the case of a DT algorithm simulation (see Table 9), 100 thousand data points would consume 3.25 joules while 614 features would consume 0.0134 joules. The RF simulation (see Table 10 for this scenario would consume 50.1 joules for 100 thousand data points and 0.182 joules for 614 features.

For a mobile recommendation system based on users' interests using an RF algorithm, 904 datapoints were used along with 1000 features [11]. 904 data points would consume 0.453 joules and 1000 features would consume 0.296 joules.

Finally, the largest scale application of machine learning found in the literature review uses an RF algorithm for the automatic detection of cash-out fraud with a training set of 131 million data points and 300 features [43]. The amount of energy consumed for 131 million data points and 300 features, would be 65.631 kilojoules and 0.0888 joules respectively.

4 DISCUSSION & LIMITATIONS

This study conducted 2 literature reviews of which 1 was non-systematic regarding the factors of energy consumption of machine learning processes. This can include biases of which are not quite clear. Therefore a systematic literature review should be conducted to remove any biases introduced due to that. Furthermore, the systematic literature review could also still be improved upon by making a wider search. Some applications were not found in the search, such as machine learning in logistics operations, as a result.

The model used for simulations is not as accurate as possible due to the omission of some factors for energy consumption. This can be further improved upon by conducting experiments regarding these factors by finding out and estimating the correlation between the factors and energy consumption. Moreover, a model regarding energy consumption while predicting (using the model) should be investigated. An interesting relation to research is whether the way a model is trained also affects its energy usage while predicting.

Based on the results of Verdecchia et al. [36], it becomes apparent that the number of datapoints and features definitely affect energy consumption of training models. Data scientists in corporations mostly have a specific data set collected by the business itself in the factory for example. It is important to consider the increase of the model's accuracy with increasing datapoints or features comparative with the energy consumption. Although the storage of said data is another point of energy consumption, using all the data is not necessary. Based on the simulations of the RF algorithm, consumption of energy can start from around 0.75 J although that is quite a small dataset in terms of number of datapoints, however the number of features used is the most out of any application identified in the literature review. Moreover, energy used can scale as high as 65.63 kJ when utilizing 131 million data points and 300 features. However, at some point the increase in number of datapoints or features increases the accuracy by a negligible amount for a huge increase in energy consumption depending on the algorithm [36].

5 FUTURE WORK

Conducting research outside of lab settings is crucial as those continuous applications of machine learning and data mining would have the most implications on energy consumption. It would be helpful to consider a case study for large scale applications such as Google's Analytics services for example. Multiple questions come to mind when investigating such a thing: How often does the model get retrained? How big is the dataset? How many features does each datapoint contain? How much energy would inference (through the trained model) consume? Performing such a case study will help bring us closer to understanding and breaking down energy consumption in real-life scenarios.

REFERENCES

[1] Christoph Boden, Andrea Spina, Tilmann Rabl, and Volker Markl. 2017. Benchmarking data flow systems for scalable machine learning. In *4th ACM SIGMOD Workshop on Algorithms and Systems for MapReduce and Beyond, BeyondMR 2017*. Association for Computing Machinery, Inc, Chicago, 1–10. https://doi.org/10.1145/3070607.3070612

[2] Alexander E I Brownlee, Jason Adair, Saemundur O Haraldsson, and John Jabbo. 2021. Exploring the Accuracy – Energy Trade-off in Machine Learning. In *2021 IEEE/ACM International Workshop on Genetic Improvement (GI)*. 11–18. https://doi.org/10.1109/GI52543.2021.00011

[3] Ye Chen, John F. Canny, and Dmitry Pavlov. 2010. Behavioral Targeting. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 4, 4 (10 2010), 31. https://doi.org/10.1145/1857947.1857949

[4] Sungwoon Choi, Jangho Lee, Min Gyu Kang, Hyeyoung Min, Yoon Seok Chang, and Sungroh Yoon. 2017. Large-scale machine learning of media outlets for understanding public reactions to nation-wide viral infection outbreaks. *Methods* 129, 1 (10 2017), 50–59. https://doi.org/10.1016/J.YMETH.2017.07.027

[5] Ronan Collobert. 2004. *Large scale machine learning.* Technical Report. Université de Paris VI.

[6] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep Neural Networks for YouTube Recommendations. In *10th ACM Conference on Recommender Systems*. ACM, New York, NY, USA, 291–198. https://doi.org/10.1145/2959100.2959190

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*. Institute of Electrical and Electronics Engineers (IEEE), Miami, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

[8] Chantat Eksombatchai, Pranav Jindal, Jerry Zitao Liu, Yuchen Liu, Rahul Sharma, Charles Sugnet, Mark Ulrich, and Jure Leskovec. 2018. Pixie: A system for recommending 3+ billion items to 200+ million users in real-time. In *World Wide Web Conference (WWW)*. Association for Computing Machinery, Inc, Lyon, 1775–1784. https://doi.org/10.1145/3178876.3186183

[9] Bella Fadida-Specktor. 2018. Preprocessing Prediction of Advanced Algorithms for Medical Imaging. *Journal of Digital Imaging* 31, 1 (2 2018), 42–50. https://doi.org/10.1007/s10278-017-9999-9

[10] Scott Fortmann-Roe. 2014. Insight Maker: A general-purpose tool for web-based modeling & simulation. *Simulation Modelling Practice and Theory* 47 (9 2014), 28–45. https://doi.org/10.1016/J.SIMPAT.2014.03.013

[11] Remo Manuel Frey, Runhua Xu, Christian Ammendola, Omar Moling, Giuseppe Giglio, and Alexander Ilic. 2017. Mobile recommendations based on interest prediction from consumer's installed apps–insights from a large-scale field study. *Information Systems* 71 (11 2017), 152–163. https://doi.org/10.1016/J.IS.2017.08.006

[12] Eva García-Martín, Crefeda Faviola Rodrigues, Graham Riley, and Håkan Grahn. 2019. Estimation of energy consumption in machine learning. *J. Parallel and Distrib. Comput.* 134 (12 2019), 75–88. https://doi.org/10.1016/j.jpdc.2019.07.007

[13] Faizan Javed and Ferosh Jacob. 2016. Data science and big data analytics at career builder. In *Big-Data Analytics and Cloud Computing: Theory, Algorithms and Applications* (1 ed.). Springer International Publishing, Derby, Chapter 6, 83–96. https://doi.org/10.1007/978-3-319-25313-8

[14] Barbara Kitchenham. 2004. *Procedures for Performing Systematic Reviews.* Technical Report. Keele University, Department of Computer Science, Keele University, UK.

[15] Heiner Lasi, Peter Fettke, Hans Georg Kemper, Thomas Feld, and Michael Hoffmann. 2014. Industry 4.0. *Business and Information Systems Engineering* 6, 4 (8 2014), 239–242. https://doi.org/10.1007/s12599-014-0334-4

[16] June Hyuck Lee, Sang Do Noh, Hyun Jung Kim, and Yong Shin Kang. 2018. Implementation of Cyber-Physical Production Systems for Quality Prediction and Operation Control in Metal Casting. *Sensors* 18, 5 (5 2018), 1428. https://doi.org/10.3390/S18051428

[17] Anna Leontjeva, Moises Goldszmidt, Yinglian Xie, Fang Yu, and Martín Abadi. 2013. Early Security Classification of Skype Users via Machine Learning. In *ACM workshop on Artificial intelligence and security*. ACM, New York, NY, USA, 35–44. https://doi.org/10.1145/2517312.2517322

[18] Shih Chieh Lin, Chang Hong Hsu, Walter Talamonti, Yunqi Zhang, Steve Oney, Jason Mars, and Lingjia Tang. 2018. ADASA: A conversational in-vehicle digital assistant for advanced driver assistance features. In *ACM Symposium on User Interface Software and Technology*, Vol. 18. Association for Computing Machinery, Inc, New York, 531–542. https://doi.org/10.1145/3242587.3242593

[19] Jelena D. Lubura, Predrag Kojić, Jelena Pavličević, Bojana Ikonić, Radovan Omorjan, and Oskar Bera. 2021. Prediction of rubber vulcanization using an artificial neural network. *Hemijska Industrija* 75, 5 (2021), 277–283. https://doi.org/10.2298/HEMIND210511026L

[20] Lucy Ellen Lwakatare, Aiswarya Raj, Ivica Crnkovic, Jan Bosch, and Helena Holmström Olsson. 2020. Large-scale machine learning systems in real-world industrial settings: A review of challenges and solutions. *Information and Software Technology* 127 (11 2020), 106368. https://doi.org/10.1016/J.INFSOF.2020.106368

[21] Jonathan Magnusson and Tor Kvernvik. 2012. Subscriber classification within telecom networks utilizing big data technologies and machine learning. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, 77–84. https://doi.org/10.1145/2351316.2351327

[22] Rashid Mijumbi, Abhaya Asthana, Markku Koivunen, Fu Haiyong, and Qinjun Zhu. 2021. Design, implementation, and evaluation of learning algorithms for dynamic real-time network monitoring. *International Journal of Network Management* 31, 4 (7 2021), e2108. https://doi.org/10.1002/NEM.2108

[23] Pieter J. Mosterman and Justyna Zander. 2016. Industry 4.0 as a Cyber-Physical System study. *Software and Systems Modeling* 15, 1 (2 2016), 17–29. https://doi.org/10.1007/s10270-015-0493-x

[24] Andrés Muñoz, David Scarlatti, and Pablo Costas. 2018. Real-time prediction of flight arrival times using surveillance information. In *12th European Conference on Software Architecture: Companion Proceedings*. ACM, New York, NY, USA, 1–4. https://doi.org/10.1145/3241403.3241434

[25] Ercan Oztemel and Samet Gursev. 2020. Literature review of Industry 4.0 and related technologies. *Journal of Intelligent Manufacturing* 31, 1 (1 2020), 127–182. https://doi.org/10.1007/s10845-018-1433-8

[26] C. Perlich, B. Dalessandro, T. Raeder, O. Stitelman, and F. Provost. 2014. Machine learning for targeted display advertising: Transfer learning in action. *Machine Learning* 95, 1 (5 2014), 103–127. https://doi.org/10.1007/s10994-013-5375-2

[27] Willem Raes, Jorik De Bruycker, and Nobby Stevens. 2021. A Cellular Approach for Large Scale, Machine Learning Based Visible Light Positioning Solutions. In *International Conference on Indoor Positioning and Indoor Navigation*. Institute of Electrical and Electronics Engineers Inc., Lloret de Mar, 1–6. https://doi.org/10.1109/IPIN51156.2021.9662610

[28] Johannes Schneider, Marcus Basalla, and Stefan Seidel. 2019. Principles of green data mining. In *Hawaii International Conference on System Sciences*, Vol. 2019-January. IEEE Computer Society, Hawaii, 2065–2074. https://doi.org/10.24251/HICSS.2019.250

[29] Roy Schwartz, Jesse Dodge, Noah A. Smith, and Oren Etzioni. 2020. Green AI. *Commun. ACM* 63, 12 (11 2020), 54–63. https://doi.org/10.1145/3381831

[30] Michael Sony and Subhash Naik. 2020. Key ingredients for evaluating Industry 4.0 readiness for organizations: a literature review. *Benchmarking: An International Journal* 27, 7 (8 2020), 2213–2232. https://doi.org/10.1108/BIJ-09-2018-0284

[31] Tao Stein, Erdong Chen, and Karan Mangla. 2011. Facebook Immune System. In *4th Workshop on Social Network Systems - SNS*. ACM Press, New York, New York, USA, 1–8. https://doi.org/10.1145/1989656.1989664

[32] Lorenzo Tieghi, Alessandro Corsini, Giovanni Delibra, and Gino Angelini. 2020. Assessment of a machine-learnt adaptive wall-function in a compressor cascade with sinusoidal leading edge. *Journal of Engineering for Gas Turbines and Power* 142, 12 (12 2020), 1–8. https://doi.org/10.1115/1.4048568

[33] Efthimios Tsiogas, Ioannis Kleitsiotis, Ioannis Kostavelis, Andreas Kargakos, Dimitris Giakoumis, Marc Bosch-Jorge, Raquel Julia Ros, R. L. Tarazon, Spyridon Likothanassis, and Dimitrios Tzovaras. 2021. Pallet detection and docking strategy for autonomous pallet truck AGV operation. In *IEEE International Conference on Intelligent Robots and Systems*. Institute of Electrical and Electronics Engineers Inc., Prague, 3444–3451. https://doi.org/10.1109/IROS51168.2021.9636270

[34] United Nations. 2015. Paris Agreement. https://unfccc.int/sites/default/files/english_paris_agreement.pdf.

[35] United Nations. 2021. Nationally determined contributions under the Paris Agreement. https://unfccc.int/sites/default/files/resource/cma2021_08_adv_1.pdf.

[36] Roberto Verdecchia, Luìs Cruz, June Sallou, Michelle Lin, James Wickenden, and Estelle Hotellier. 2022. Data-Centric Green AI: An Exploratory Empirical Study.

[37] Jizhe Wang, Pipei Huang, Huan Zhao, Zhibo Zhang, Binqiang Zhao, and Dik Lun Lee. 2018. Billion-scale Commodity Embedding for E-commerce Recommendation in Alibaba. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 18 (3 2018), 839–848. https://doi.org/10.48550/arxiv.1803.02349

[38] Meng Wang, Weijie Fu, Xiangnan He, Shijie Hao, and Xindong Wu. 2020. A Survey on Large-scale Machine Learning. https://doi.org/10.48550/ARXIV.2008.03911

[39] Shuhao Wang, Cancheng Liu, Xiang Gao, Hongtao Qu, and Wei Xu. 2017. Session-Based Fraud Detection in Online E-Commerce Transactions Using Recurrent Neural Networks. *ECML PKDD Machine Learning and Knowledge Discovery in Databases* 10536 LNAI (2017), 241–252. https://doi.org/10.1007/978-3-319-71273-4

[40] Claes Wohlin. 2014. Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering (EASE '14)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/2601248.2601268

[41] Shuang Hong Yang, Alek Kolcz, Andy Schlaikjer, and Pankaj Gupta. 2014. Large-scale high-precision topic modeling on twitter. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, New York, 1907–1916. https://doi.org/10.1145/2623330.2623336

[42] Jinhui Yuan, Fei Gao, Qirong Ho, Wei Dai, Jinliang Wei, Xun Zheng, Eric P. Xing, Tie Yan Liu, and Wei Ying Ma. 2015. LightLDA: Big topic models on Modest Computer Clusters. In *24th International Conference on World Wide Web*. Association for Computing Machinery, Inc, Republic and Canton of Geneva, 1351–1361. https://doi.org/10.1145/2736277.2741115

[43] Ya-lin Zhang, Jun Zhou, Ziqi Liu, Zhiqiang Zhang, Chaochao Chen, Y-l Zhang, J Zhou, L Li, Z Liu, Z Zhang, C Chen, X Li, Y Qi, W Zheng, J Feng, M Li, Z-h Zhou, Ya-Lin Zhang, Wenhao Zheng, Ji Feng, Longfei Li, Ming Li, Xiaolong Li, Yuan Qi, and Zhi-Hua Zhou. 2019. Distributed Deep Forest and its Application to Automatic Detection of Cash-Out Fraud. *ACM Transactions on Intelligent Systems and Technology (TIST)* 10, 5 (9 2019), 1–19. https://doi.org/10.1145/3342241

[44] Zhenhua Zhang, Leon Stenneth, Ram Marappan, Zaba Sebastian, Philip S Yu, Z Zhang, L Stenneth, R Marappan, Z Sebastian, and P Yu. 2018. Insert Beyond the traffic sign recognition: constructing an auto-pilot map for autonomous vehicles. In *26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, New York, NY, USA, 468–471. https://doi.org/10.1145/3274895.3274951

[45] Lu Zheng, Zhao Tan, Kun Han, and Ren Mao. 2018. Collaborative Multi-modal deep learning for the personalized product retrieval in Facebook Marketplace.

https://doi.org/10.48550/ARXIV.1805.12312

[46] Hao Zhou, Hong feng Chai, and Mao lin Qiu. 2019. Fraud detection within bankcard enrollment on mobile device based payment using machine learning. *Frontiers of Information Technology & Electronic Engineering 2018 19:12* 19, 12 (1 2019), 1537–1545. https://doi.org/10.1631/FITEE.1800580

[47] Dimitrios Zissis. 2018. Intelligent security on the edge of the cloud. *2017 International Conference on Engineering, Technology and Innovation: Engineering, Technology and Innovation Management Beyond 2020: New Challenges, New Approaches, ICE/ITMC 2017 - Proceedings* 2018-January (2 2018), 1066–1070. https://doi.org/10.1109/ICE.2017.8279999

A SIMULATION TABLE RESULTS

This section shows the results of the extrapolation simulations done through Insight Maker.

A.1 SVM Simulation

Table 7. SVM Simulation Results

Datapoints (#)	DEC (J)	Features (#)	FEC (J)	Total Energy (J)
1,600,000	198.4	240	0.022824	198.422824
1,700,000	210.8	255	0.0242505	210.824251
1,800,000	223.2	270	0.025677	223.225677
1,900,000	235.6	285	0.0271035	235.627104
2,000,000	248	300	0.02853	248.02853

A.2 KNN Simulation

Table 8. KNN Simulation Results

Datapoints (#)	DEC (J)	Features (#)	FEC (J)	Total Energy (J)
1,600,000	6	240	0.0004944	6.0004944
1,700,000	6.375	255	0.0005253	6.3755253
1,800,000	6.75	270	0.0005562	6.7505562
1,900,000	7.124	285	0.0005871	7.1255871
2,000,000	7.5	300	0.000618	7.500618

A.3 DT Simulation

Table 9. DT Simulation Results

Datapoints (#)	DEC (J)	Features (#)	FEC (J)	Total Energy (J)
80,000	2.6	512	0.0112128	2.6112128
85,000	2.7625	544	0.0119136	2.7744136
90,000	2.925	576	0.0126144	2.9376144
95,000	3.0875	608	0.0133152	73.1008152
100,000	3.25	640	0.014016	3.264016

A.4 RF Simulations

Table 10. RF Simulation Results

Datapoints (#)	DEC (J)	Features (#)	FEC (J)	Total Energy (J)
904	0.452904	1000	0.296	0.748904
...	...	...	...	...
100,000	50.1	614	0.181744	2.9376144
...	...	...	...	...
131,000,000	65,631	300	0.0888	65,631.0888