AWS Cloud: How Web Applications Can Improve Product Usability

Actors Which Inform Uses of Machine Learning Models for Sustainability at Amazon Compared to Academia

A Thesis Prospectus

In STS 4500

Presented to

The Faculty of the

School of Engineering and Applied Science

University of Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science in Computer Science

By

Aidan Ricci

October 11th, 2022

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISORS

Joshua Earle, Department of Engineering and Society

Kevin Sullivan, Computer Science

Introduction

Over recent years, sustainability has become an increasingly important piece of large organizations and a major area of research in academia. However, since sustainability is a relatively new field, new methods need to be utilized in order to monitor and increase our conservation efforts. One example of a method which can be employed to further these efforts is machine learning. Machine learning can help a computer learn how to increase sustainability in a number of different ways, but the aim of this project is to determine what actors are involved in the decision to incorporate new machine learning models or techniques at Amazon. This project will also incorporate a comparison in the ways that academia and Amazon use machine learning to move sustainability goals

For the technical project of my prospectus I built a website to increase the usability of Amazon's AWS Genomics CLI and my STS research question is: Actor Network Theory and the Creation of Machine Learning Models for Sustainability at Amazon. My technical project was built as a part of my internship at Amazon this past summer. As a result, it never had an official research question. However, the project was important because it would allow more genomics researchers to easily use the cloud for running DNA Sequencing workloads. This would allow researchers to move away from on-premises systems and allow them to leverage the benefits of cloud services. My STS research question is important because historically corporations have failed to be sustainable, preferring to maximize profits over helping humans to live into future generations. However, in current times, the largest companies claim to lead in sustainability. As a result, it would be interesting to know exactly what a corporate leader in sustainability, like Amazon, has done using machine learning to become a leader in sustainability, and how their uses of machine learning to achieve sustainability goals compares to the research available in

academia. Particularly, identifying the actor network involved in the addition of sustainability models at Amazon could help us to understand how we can enact similar sustainability efforts at other companies. Finally, understanding how Amazon's approach to using these models differs from academia could help us understand how Amazon makes sustainability work with business objectives, which provide the incentive to increase sustainability. My technical project and STS project are not connected, besides the fact that both involve Amazon. Over the course of this prospectus there will be a portion on the objectives and result of the technical project, followed by the topic, objectives, and roadmap of the STS project, and finally a piece on the key texts found in my bibliography and why I used them.

Technical Project

How can a genomics researcher conduct research without needing to buy and set up high performance computing hardware? The best solution to this problem is cloud services, which can provide the same functionality as on-premises hardware without any of the upfront cost or set up requirements. Despite services being available in the cloud today, genomics researchers do not widely use the cloud to run their DNA/RNA sequencing workloads. This happens for a number of reasons, but one contributing factor is the usability and technical knowledge requirements of some of these services. Therefore, the purpose of my summer project was to offer a service which could bring new researchers into the cloud for their genomics workloads by reducing the amount of technical knowledge needed in order to begin. This meant taking the existing Genomics CLI, which required a lot of technical knowledge (Writing YAML files, using a command line tool, knowing cloud services, etc.) and turning it into a web app to be used as a management console. This would allow users to jump straight into genomics workloads in the cloud, with little prior knowledge and set up.

STS Project

For the STS project, my research question is: What Actors Meaningfully Impact the

Creation of Machine Learning Models for Sustainability at Amazon. The project will explore

what actors are important in pushing the use of machine learning technology at Amazon for

reaching sustainability objectives. This means that I will be examining the actor network

involved in pushing the use of machine learning models for sustainability at Amazon. The topic

is important because it can help us understand the actors which help Amazon to be a

sustainability leader, allowing us to understand how the same type of change can be enacted at

other companies. The framework focused on in this project will be actor-network theory and the

research will employ a historical approach of finding, reading, and synthesizing previous

literature.

The relevant social groups that I would like to focus on in this project are Amazon hiring,

Amazon's sustainability group (Barr, 2022), Amazon consumers, Amazon upper management

(C-Suite, sustainability managers), and researchers in academia. I identified these groups because

I believe they are the most likely to directly impact Amazon sustainability in their day to day

efforts. Through my experience at Amazon and the research I did for this prospectus, I have

recognized that much of Amazon's sustainability efforts come from the culture developed at the

highest levels of the company and it is perpetuated by the employee groups and hiring decisions

made at the lower levels. As a result, I believe Amazon hiring principles, upper management,

and their employee groups are the ones who guide the overall sustainability objectives of the

company from within the company. The other two groups I decided upon for different reasons. I

decided to add Academia as a potential actor because it is the source of an enormous amount of machine learning research. I'm curious to see whether this research is an important actor at Amazon in enacting sustainability changes because it could be useful in identifying the difference in goals between sustainability research in academia and sustainability action within these companies. It could also show whether all the academic research is resulting in true action in the places where it could be useful. The final factor I chose is Amazon's consumers, because although they will not know individual models for machine learning, consumers could be applying pressure to Amazon in a unique way which makes them focus more on sustainability objectives than other companies. This is important to understand because if this is the case it means that a grassroots movement could cause similar changes at other companies. However, I also decided to leave out a couple of potentially important social groups. One which is likely relevant is the general group of all Amazon employees. I left these out because I believe that they would have less individual impact than employee groups or management, even though I do believe they have a say. Another interesting actor to take into account could be government intervention, since Amazon is such a major corporation. There may be policy or law related factors which impact Amazon in using machine learning for sustainability.

Actor/Network Theory Framework:

Actor network theory is a good framework for this research because this project focuses on the stakeholders needed to produce the machine learning models which can improve sustainability. Ideally, my project will be able to find the actors who have fostered a sustainable culture at Amazon so that their efforts can be recreated at a company. I used a historical approach for this research because ethnography would be difficult since most employees likely have NDAs and wouldn't be able to talk about this topic. In this network I hope to focus on the major people

within the company, the company culture, their consumers, and finally academic research as a comparison to see if it has impacted Amazon's own advancements. I will research this by comparing a small set of uses for machine learning in sustainability both in academia and in Amazon. This will help me determine the impact of academia on Amazon. I will then use sources like hiring principles, hiring pages, and the sustainability report to see how company hiring culture and sustainability becomes an actor in this network. Finally, I will look into the goals of individual members of senior management to see how they personally push Amazon's sustainability efforts.

Why is it important?:

Researching the important actors which need to be involved in the adoption of new machine learning techniques for sustainability is important for ensuring other companies can enact the change to make this world more sustainable. It could make these changes easier to implement in the future, allowing companies to excel further in their sustainability objectives and move our world closer to where we need to be without the government reform which has been failing to be implemented for decades.

Key Texts

The Amazon 2021 Sustainability Report (Amazon.com inc., 2022) is the entire sustainability report released by Amazon this year . It documents all of Amazons sustainability efforts from the progress they have made towards their climate pledge to their current carbon footprint, to how they are changing their packaging to be more sustainable. This 100 page report is important to my project because it is an in depth look into all of Amazon's largest sustainability initiatives and how effective they have been. It allows me to not only see how

Amazon is faring in their climate pledge, but also all the important ways they have been trying to hit their goals. Being able to see all these initiatives can allow me to better understand how Amazon is trying to be sustainable and how/whether Machine Learning is being used to fulfill these goals. It also gives me insight into how their decisions are being made and potentially some of the actors involved in these decisions.

Measuring sustainability through ecological sustainability and human sustainability: A machine learning approach (Nilashi, 2019) was written by a number of academics at universities throughout Asia. It cites a number of articles from the same area of the world focused on using machine learning to assess sustainability performance. They leave out the other half of stakeholders, the companies, for the most part in this paper. The value of this paper to these two projects is that it represents a large area of research in machine learning sustainability, the development of assessments and measures. They attempt to use fuzzy clustering and supervised machine learning to give country sustainability assessments. This used a large number of different pieces of data to show results and represented a potentially valuable approach to measuring sustainability. It is possible that larger companies could emulate this approach.

A supervised machine learning classification framework for clothing products' sustainability (Satinet, 2022) is an academic journal focused on determining how to identify environmentally sustainable clothing. The authors realize that "there are a variety of confusing sustainability certifications and few labels capturing the overall environmental impact of products"(Satinet, Fouss, 1), which they believe is a result of this analysis being costly. As a result the paper explores the use of supervised machine learning tools to develop a model to assess the environmental sustainability of clothing by using existing assessments. Their models

found success in their random forest algorithm which had an average accuracy of 91% over the five folds. It could be used to quickly provide interested consumers with productlevel sustainability information, or even to develop a unique and all-inclusive environmental label. This represents a potential use of machine learning which is currently not in use by large companies, like Amazon, today.

Do online comments affect environmental management? identifying factors related to environmental management and sustainability of Hotels (Saura, 2018) is a study aimed to identify the key indicators related to environmental management and sustainability of hotels as perceived by travelers during their trips. They analyzed the trip advisor reviews on sustainability to determine what consumers most cared about in sustainability. The goal of this was to help guide hotels in how to improve their services and their sustainability policies to better respect the environment and line with their customer wants. Although this paper is limited in its sample size, it does represent another possible way that large companies can guide their sustainability objectives by giving their customers a larger voice.

An exploratory DEA and Machine Learning Framework for the evaluation and analysis of sustainability composite indicators in the EU. (Tsaples, 2022) is an academic journal which proposes an alternative method for measuring sustainability which employs machine learning to increase the realism of the results received in order to increase trust in the measurement. The method is used to calculate the sustainability of the 28 EU countries in this paper, so it has not been applied to businesses, however the benefit to the creation of a model like this for business could be great. Giving the public a standardized, reliable measure of the sustainability of the companies they buy products from may force companies to fight over their sustainability

measures in order to retain environmentally minded customers. This represents another approach to improving sustainability at large companies.

The second source (Barr, 2022) is a hiring page for why people should choose to work at Amazon because of sustainability. It outlines not only their goals for becoming more sustainable, but also primary examples for what they are currently doing to become more sustainable, like using electric transportation vehicles and low carbon aircraft. Additionally it touches upon Amazon's important "Sustainability Ambassadors" program, which allows employees to provide their own inputs on how Amazon can become more sustainable, whilst also giving them the opportunity to work towards these goals. This is important to my project because it gives the argument for how any employee can become involved in the decision making process for sustainability at Amazon and an example of how you can increase your voice within the company for these objectives.

Reduce food waste to improve sustainability and financial results in retail with Amazon Forecast (Cheng, 2008) is a blog post by Amazon which provides a good example of where Amazon ML solutions are being used to improve sustainability at another company. In particular, this blog post talks about how ML is being used at one of Greater China's top convenience store chains to reduce food waste, allowing the chain to save money and hit sustainability goals. This is done by forecasting the demand of different products sold by the chain. This is important to my project because it represents a core differentiator in the use of ML by Amazon and their clients as opposed to academia. In academia the use seems to be more specific and oriented towards general benefits, whereas Amazon here has managed to become more sustainable by helping companies save money and being valuable to the customer.

Forecasting Amazon Rain-forest deforestation using a hybrid machine learning model (Dominguez, 2022) is an academic paper which documents the use of a hybrid machine learning model to forecast the deforestation of the Amazon rainforest, a major concern for environmental sustainability. In this paper they used geographical data from Brazilian municipalities and a 20 year time series of data on annual deforestation to create a machine learning model that can predict future deforestation trends. This is important to my paper because it represents an academic use of ML to document trends which combat sustainability efforts in likely the most important area of the world for climate initiatives. The general documentation for informing academia and the public of sustainability concerns shows a difference in the uses of ML for academia and the Amazon corporation because it does not fulfill any business goals.

Automating Wind Farm Maintenance Using Drones and AI (Gozluklu, 2008) is another AWS blog post which documents how AWS' machine learning and internet of things services are being used to automate the expensive, high-risk task of maintaining the wind farms which create sustainable energy. This blog post is useful to my project because it gives a new perspective to view ML sustainability through. It is a new lens because the post documents a machine learning solution which supports environmental initiatives rather than directly causing sustainability, since what the solution does is reduce the cost and risk of repairing our sustainable energy rather than causing sustainability through the ML results. The other interesting piece of this blog is that it documents exactly which AWS services were used for each piece of the solution and it gives an overview of the entire cloud architecture so people can understand better how it works.

Using the AWS Cloud to restore ecosystems around the world (Wanser, 2008) is another AWS blog post on how AWS services are being employed at client companies to improve the environment. In this case the blog post is about how a company named Dendra systems is using

AWS and big data assess and restore at-risk ecosystems. They start by talking about how we need to restore and expand native forests to capture greenhouse gases rather than simply planting trees. It then speaks about how Dendra systems is using AWS AI to identify animals and slight differences in nature that are nearly imperceptible to humans in order to gather insights which can optimize the health of an ecosystem. It can even supposedly make automated decisions like "aerial planting" seeds into an ecosystem. This is important to my paper for two reasons. The first is that it further exemplifies how Amazon optimizes sustainability for others as well as themselves by providing the tools and hardware in AWS for others to use. It also shows an example of what one might think more of a "typical" use of Machine Learning for sustainability objectives.

References

Amazon.com Inc.. 2022. *Amazon's 2021 Sustainability Report.*

Barr, C., & Claude, J. (2022). *Let's save the Amazon: Why we must protect our planet*. Amazon.

Retrieved December 8, 2022, from

https://hiring.amazon.com/why-amazon/sustainability#/

Cheng, J. (2008). *Reduce food waste to improve sustainability and financial results in retail with*

Amazon Forecast. Amazon. Retrieved December 8, 2022, from

https://aws.amazon.com/blogs/machine-learning/reduce-food-waste-to-improve-sustainab

ility-and-financial-results-in-retail-with-amazon-forecast/

Dominguez, D., del Villar, L. de, Pantoja, O., & González-Rodríguez, M. (2022). Forecasting

Amazon Rain-forest deforestation using a hybrid machine learning model. *Sustainability*,

14(2), 691. https://doi.org/10.3390/su14020691

Gozluklu, B. (2008). *Automating Wind Farm Maintenance Using Drones and AI*. Amazon.

Retrieved December 8, 2022, from

https://aws.amazon.com/blogs/industries/automating-wind-farm-maintenance-using-dron

es-and-ai/

Nilashi, M., Rupani, P. F., Rupani, M. M., Kamyab, H., Shao, W., Ahmadi, H., Rashid, T. A., &

Aljojo, N. (2019). Measuring sustainability through ecological sustainability and human

sustainability: A machine learning approach. *Journal of Cleaner Production*, *240*,

118162. https://doi.org/10.1016/j.jclepro.2019.118162

Satinet, C., & Fouss, F. (2022). A supervised machine learning classification framework for clothing products' sustainability. *Sustainability*, *14*(3), 1334. https://doi.org/10.3390/su14031334

Saura, J., Reyes-Menendez, A., & Alvarez-Alonso, C. (2018). Do online comments affect environmental management? identifying factors related to environmental management and sustainability of Hotels. *Sustainability*, *10*(9), 3016. https://doi.org/10.3390/su10093016

Tsaples, G., Papathanasiou, J., & Georgiou, A. C. (2022). An exploratory DEA and Machine Learning Framework for the evaluation and analysis of sustainability composite indicators in the EU. *Mathematics*, *10*(13), 2277. https://doi.org/10.3390/math10132277

Wanser, K. (2008). *Using the AWS Cloud to restore ecosystems around the world*. Amazon. Retrieved December 8, 2022, from https://aws.amazon.com/blogs/publicsector/using-aws-cloud-restore-ecosystems-around-world/