Noor Benny

Professor Haffa

English 2

9 December 2020

Saving the Environment with Data Science

The global population count is at 7.7 billion people and growing; however, humans are using natural resources and polluting the earth at an unsustainable rate. In order to be able to continue supporting life on planet earth, humans must find sustainable ways to slow down the rate at which resources are used. One way that researchers can find answers to this issue is by looking at the mass amounts of information gathered by the latest and most advanced technologies. By collecting and utilizing big data, data scientists can work with environmental specialists to solve environmental issues on a local, national, and global scale.

Data science is a broad field of study centered around the collection and organization of data. More specifically, data science can be defined as "a set of fundamental principles that support and guide the principled extraction of information and knowledge from data" (Fawcett). Data science is most closely related to data mining and uses data mining principles when working with information. Data mining is the exploration of structures and patterns in large sets of information (Hand). Data mining can be split into two categories: model building, or data analysis, and data detection, or information filtering (Hand). Similarly to data science, most work in data mining "has been computational, with an emphasis on algorithms" (Hand). Additionally, data collection and analytics is an extremely new field of research, only founded after the

creation and implementation of technologies that are capable of gathering data at an efficient rate. In summary, data science relies on technology and algorithms in order to gather and analyze patterns in information.

Despite being an extremely new field of study, the practical applications of data analysis have been utilized by a broad range of researchers. For example, medical specialists currently use data collection to personalize healthcare recommendations (Rice). On a much larger scale, data scientists are able to use information to optimize shipping routes in real-time (Rice). Using data science, Google went as far as to discover a tool, LYNA, that can identify cancerous tumors that can be difficult for the eye to see (Rice). The tool was able to "accurately identify metastatic cancer 99 percent of the time using its machine-learning algorithm" (Rice). From commercial to medical, data science has been able to shape almost every industry; however, the question still remains unanswered—how can researchers use data science to approach environmental issues on a local, national, and global scale?

Before tackling the question, researchers must identify environmental issues that exist at each level. At a local level, scientists can turn to the water crisis in Flint, Michigan. When the city of Flint decided to switch their water source from Detroit's system to the Flint River, experts failed to adequately treat and test the water before distributing it to local homes and businesses (Denchak). The mistake resulted in "foul-smelling, discolored, and off-tasting water [being] piped into Flint homes for eighteen months [and] was causing skin rashes, hair loss, and itchy skin" because the water was contaminated by pollution from local industries (Denchak). On the other side of the country, California faced the effects of global warming on a local scale. Drier,

warmer conditions, which are a direct result of climate change, increased droughts and prolonged the wildfire season. On a national scale, the Amazon Rainforest in Brazil has been facing deforestation at an alarming rate. The rampant deforestation has caused species to slowly go extinct due to a loss of habitat. Most, if not all, of the aforementioned issues contribute to environmental problems that exist on a global scale. One of the most well known global environmental issues is global warming. The National Aeronautic and Space Association defines global warming as "the long-term heating of Earth's climate system observed since the pre-industrial period (between 1850 and 1900) due to human activities, primarily fossil fuel burning, which increases heat-trapping greenhouse gas levels in Earth's atmosphere" (NASA). The rise in the temperature of the climate is causing the Arctic sea ice to diminish which directly correlates to a rise in the global sea level. In order to approach environmental issues using data science, researchers must identify the problems at each level.

Currently, specialists use a myriad of approaches to solve environmental problems; however, they are often faced with challenges when approaching these issues. Firstly, researchers need mass amounts of information in order to use to solve the issues at hand. Although technology has seen huge advancements in the recent decades, many researchers cannot begin to solve the environmental issues due to a shortage of information. Many scientists do not have algorithmic and organized ways of storing and analyzing information, which makes it difficult to come up with small scale solutions. Instead, environmental scientists attempt to enact change on national levels by pressuring law and policy makers (Student Scholarships). For example, the Green New Deal, "the most high-profile piece of environmental legislation," outlines a plan to

address climate change and economic inequality “with the goal of getting to a one-hundred percent clean energy economy by 2050” (Karaim). Large scale solutions can be helpful; however, information from consumer patterns that can be collected using big data is inaccessible, therefore keeping environmental scientists from delving into human behavior and its correlation to environmental issues. Furthermore, scientists are working on genetically modifying plant genes in order to slow down climate change. In the European Union, the government supported research “aimed at developing drought-tolerant cereal [that] floods, drought and other hazards such as wildfires are some of the problems that climate change will exacerbate, especially in the poorest regions” (OpenMind). Environmental specialists continue to be faced with challenges when approaching environmental problems.

One answer to the shortage of information is the implementation of data science. Several aspects of the field lend themselves to solving complex problems including the environmental issues at hand. Firstly, the manner by which data is collected allows scientists to solve environmental problems more efficiently. When data is collected algorithmically, it can be stored in an organized way that is easy to find and access. More importantly, the manner by which data is organized and stored allows scientists to solve environmental problems more efficiently. After being collected, data can be organized into four categories: volume, or the size or the data set, variety, or the multiple types of data gathered, velocity, or how fast the data can be accessed, and veracity, or the algorithms required to ensure the data quality and accuracy (Rominger). Each of the datatypes has its own functions and benefits. Volume allows for a large amount of data to be gathered and allows data scientists to analyze almost all possible sides to an issue (Mason). Next,

variety structures the information in different ways and allows the information to be presented in different ways for clarity and understanding (Mason). Moreover, velocity allows data to be gathered rapidly in order to react in a timely manner (Mason). Finally, veracity ensures that the data is verifiable and checks for accuracy and integrity (Mason). The four categories, often referred to as the "four v's of big data", are important because as a data set grows larger, more potential patterns can be identified (Faghmous). The patterns can be used to find solutions to issues in the most efficient ways possible. The four data types work together to present scientists with solutions to issues, including about the environment.

Researchers have successfully begun to use big data to solve issues in the environment. For example, climate scientists collect and analyze information through Geographic Information System databases (Faghmous). A Geographic Information System, also referred to as a GIS, database is "a powerful tool that helps scientists conceptualize complex environmental conditions and manage large amounts of information [and] can assimilate information from multiple sources and time periods about a location or series of locations into a database structure" (Faghmous). In simpler terms, GIS databases help manage large influxes of data that is collected from the environment and the data is used to predict environmental trends. The trends in the climate and the environment can be predicted using patterns which are often presented using visual models. More practically, "creating different models for different scenarios [can help scientists] see how likely it is that climate change will be exacerbated by various policy decisions and lifestyle choices at the macro level" (Ryan). More generally, patterns in data can be used to get to the source of several environmental issues including the correlation of deforestation with

the overconsumption of paper products on a global level. On a local level, researchers can find ways to reduce over-consumption of water in California by linking water consumption to human patterns. Scientists can even use data analysis to reduce water consumption and use genetic modification to lessen the agricultural industry's impact on the environment. The patterns could also be used to link human transportation patterns with fossil fuel emissions and air pollution. Addressing human contribution to climate change through data science can be a step in the right direction to saving planet earth. By using data science, researchers can use large amounts of organized information in order to solve environmental issues on a local, national, and global scale.

Although many advantages exist to using big data, scientists are faced with challenges and concerns when extracting mass amounts of information. Humans have the largest negative impact on the environment, so scientists must analyze consumer patterns in order to explore trends between human behavior and environmental issues. However, many people have privacy concerns when large amounts of information are extracted from their lives. Generally, individuals receive notice that their information is being collected ahead of time; however, many individuals are misinformed about the extent to which their information can be used. The information can fall "in the hands of marketers, financial institutions, employers and government, can affect everything from relationships to getting a job, and from qualifying for a loan to even getting on a plane" (Armerding). Additionally, limitations also exist when extracting data using different methods. Extracting raw information can be easy, but researchers often have trouble accessing or transferring user-level data (Yamaguchi). Moreover, user-level algorithms

have difficulty answering the “why” of different issues (Yamaguchi). Other than limitations, extraction of data also has ethical concerns. When large amounts of information are collected from an individual, their identity can be compromised. Although data analysis and algorithms can be used to cater to a person’s likes and dislikes, analyzing data from an individual permits institutions to regulate and determine individual characteristics and attributes at an alarming rate (Price). As a result, “civil liberties advocates want to control the use of big data, and others think companies should pay to use people's online information” (Price). Although data science has its benefits, scientists continue to be faced with challenges and concerns when extracting mass amounts of data from consumers.

Humankind harms the earth more than any other species on this planet; therefore, humans have the responsibility to find solutions to environmental issues to reverse its effects on the earth. In order to sustain life, researchers must find sustainable ways to live in order to slow down the rate of consumption. Data science, a new technological field of study, can help solve the environmental issues at hand by analyzing mass amounts of information from around the world. Researchers already use data science to approach a myriad of issues; however, the field is yet to be applied to environmental conservation on a larger scale. Although limitations due to privacy and ethical concerns exist, the benefits of data extraction in order to save the environment outweigh the individual consequences. Through the use of big data, data scientists and environmental specialists can work together in order to solve environmental issues on a local, national, and global scale.

Works Cited

Armerding, Taylor. "The 5 Worst Big Data Privacy Risks (and How to Guard against Them)." CSO Online, CSO, 14 July 2017, www.csoonline.com/article/2855641/the-5-worst-big-data-privacy-risks-and-how-to-gua d-against-them.html.

Denchak, Melissa. "Flint Water Crisis: Everything You Need to Know." NRDC, 8 Nov. 2018, www.nrdc.org/stories/flint-water-crisis-everything-you-need-know.

"Environmental Scientists and Specialists - What Do Environmental Scientists and Specialists Do?" StudentScholarships.org, Student Scholarships, 2020, studentscholarships.org/salary/370/environmental_scientists_and_specialists.php.

Fawcett, Foster Provost and Tom, et al. "Data Science and Its Relationship to Big Data and Data-Driven Decision Making." *Mary Ann Liebert, Inc., Publishers*, 13 Feb. 2013, www.liebertpub.com/doi/full/10.1089/big.2013.1508.

Hand, David J, and Niall M Adams. "Data Mining." *Wiley Online Library*, American Cancer Society, 22 June 2015, onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat06466.pub2.

James H. Faghmous and Vipin Kumar, et al. "A Big Data Guide to Understanding Climate Change: The Case for Theory-Guided Data Science." *Mary Ann Liebert, Inc., Publishers*, 15 Sept. 2014, www.liebertpub.com/doi/full/10.1089/big.2014.0026.

Karaim, Reed. "Fuel Efficient Standards." *CQ Reseacher*, CQ Researcher, 19 June 2020, library.cqpress.com.libproxy.mpc.edu/cqresearcher/document.php?id.

Kh, Ryan. "Can Big Data Help Save The Environment? SmartData Collective." *SmartData Collective*, 21 Nov. 2020,

www.smartdatacollective.com/can-big-data-help-save-environment/.

Mason, Matthew G. “Big Data: Explaining Its Uses to Environmental Sciences.” *EnvironmentalScience.org*, www.environmentalscience.org/data-science-big-data.

OpenMind. “Technologies That Can Save the Environment.” *OpenMind*, 21 Apr. 2016, www.bbvaopenmind.com/en/science/environment/technologies-that-can-save-the-environment/.

“Overview: Weather, Global Warming and Climate Change.” NASA, NASA, 28 Aug. 2019, climate.nasa.gov/resources/global-warming-vs-climate-change/.

Price, Tom. “Should the Use of Personal Information Be Restricted?” *CQ Researcher*, CQ Researcher, 25 Oct. 2013, library.cqpress.com.libproxy.mpc.edu/cqresearcher/document.php?id=cqresrre2013102500.

Rice, Mae. “17 Data Science Applications & Examples.” Built In, 23 July 2019, builtin.com/data-science/data-science-applications-examples.

Rominger, Jeff, and Shingo Ikeda. “The Role of Big Data in Solving Environmental Problems.” *Gradient*, 2015, gradientcorp.com/pdfs/newsletter/Trends64.pdf.

Tsui, Jenna. “Five Biggest Environmental Issues Affecting the U.S.” *Environmental Protection*, EP Online, 24 Feb. 2020, eponline.com/Articles/2020/02/24/Five-Biggest-Environmental-Issues-Affecting-the-US.

Yamaguchi, Kohki. “7 Limitations Of Big Data In Marketing Analytics.” *Marketing Land*, 13 Feb. 2015, marketingland.com/7-limitations-big-data-marketing-analytics-117998.aspx?Page=4.