Utilizing Machine Learning to Predict and Analyze the Locations and Movements of Invasive Burmese Pythons in the Florida Everglades
(Technical Topic)

Providing a Data-Driven, Unified Framework for Public and Private Entities to Combat the Threat of Invasive Species
(STS Topic)

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
Samuel Veliveli
April 2, 2024

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.

Signature: _____

Advisors

Kathryn A. Neeley, Department of Engineering and Society

Introduction

The Florida Everglades is a rich, biodiverse ecosystem of about 1.5 million acres of wetlands in the southern tip of Florida. The Everglades and its habitats support numerous endangered species as well as more than 360 bird species, 300 types of fresh and saltwater fish, 40 different mammal species, and 50 species of reptiles. Everglades National Park is protected by the National Park Service, making it the only federally protected subtropical wilderness area in North America. Parts of the Everglades consist of areas that are close to urban activities - unfortunately, this has caused this keystone ecosystem to be prey to the exotic pet trade. More specifically, exotic pet owners will abandon their pets in the Everglades once it no longer becomes convenient to have them. As a result, exotic fishes have been devouring native fish species and nonnative plants often shade out indigenous plants. Of the most devastating of these invasive species, none have had the impact that the Burmese Python has had on the Everglades. Almost perfectly suited for the Everglades, with coloring that blends in with the environment and no natural predators, they have eaten almost everything in their path. This has caused many species in the Everglades, particularly mammals to decline very sharply in number (Fort Collins Science Center, 3).

If species like the Burmese Python are not effectively dealt with, they could destroy an ecosystem that is home to many species. Since so many species are connected to the Everglades, this could cause catastrophic effects in the food web and ultimately affect us as humans. To stabilize the situation, I am proposing a combined technical and sociotechnical approach to resolution. I plan on utilizing machine learning algorithms to identify geographic patterns in the pythons' movements using public data, and displaying that data on a publicly available server,

while also using machine learning as a unifying tool between the public and private domains of conservation.

Technical Topic:

Utilizing Machine Learning to Predict and Analyze the Locations and Movements of Invasive Burmese Pythons in the Florida Everglades

The program I will be using to make my potential deliverable is called Tableau. Tableau is a well-known data visualization and analytics software. Tableau has been used for numerous ecological problems as well as for displaying geospatial data. The great thing about Tableau is that it provides free servers for users to upload their dashboards to make their content easily available and accessible. In terms of machine learning, I will be utilizing a technique called K-means clustering. As shown in Figure 1, K-means clustering takes a set of data points and calculates a type of mean called a centroid. Based on the point's proximity to the nearest centroid, the points are reassigned to groups. This type of clustering is especially useful in identifying patterns in geospatial data. Additionally, K-means is well-known in the machine learning community and can be explained easily. It is not so complex that a researcher could not explain the model, but effective enough that it can help derive potential insights from data. Combining machine learning capabilities with Tableau's data analytics and visualization will provide a robust and powerful tool for conservation as shown by other research (Zhang, 1). The resulting dashboard will potentially allow for hidden insights and patterns that have never been seen before due to the lack of modern technology in the field of conservation ecology. Current conservation methods (in this case manually removing pythons based on their location) simply take too much time and are not a good use of resources. By utilizing the latest in machine

learning, we can allocate our resources more effectively and efficiently, by seeing areas of high density of the pythons and where they are likely to frequent.

How K-Means Clusters Raw Data

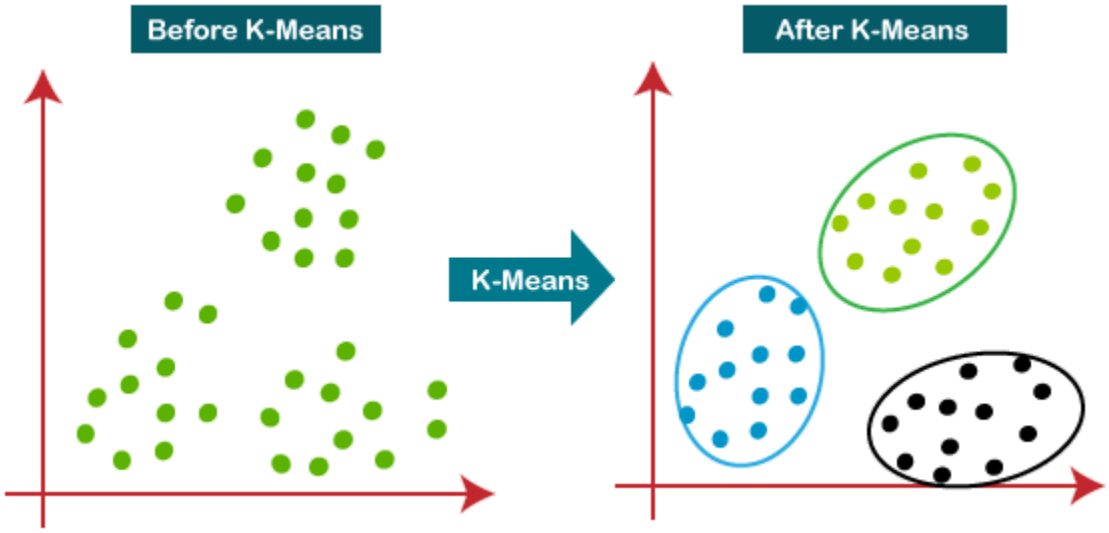

Figure 1. Sample Data with K Means Applied. (Setruck 1).

My research will build upon existing knowledge of tracking Burmese Pythons. Some apps have citizens manually enter locations where they have seen Pythons, and there is a larger movement in the ecological community to utilize techniques such as artificial intelligence and machine learning (Fang, 227). As the conservation field has slowly become more modern, there are researchers trying to integrate machine learning for different ecological purposes such as mapping invasive plant species (Singh, 1). Additionally, some of these researchers are focusing on the Everglades with one focusing on remote monitoring of species (Zhang, 1). However, I have found very little in using machine learning to monitor and predict the locations of pythons and invasive species in general. The main challenge I anticipate is finding accurate data

regarding geospatial data for the python. It will take some time to cross-validate datasets and make sure I am getting quality data. Once this data is gathered, the dashboard will be relatively simple to make.

Taking a brief moment to introduce and address the sociotechnical side of this issue, the python problem in the Florida Everglades is also important due to its symbolic nature representing cross-collaborative conservation efforts. Numerous government agencies, university scientists, and also regular citizens have collaborated to try and fix this problem. (Fort Collins Science Center, 3). If the python problem can be effectively mitigated, this could provide a framework for how the public and private sectors can integrate to stop invasive species as a whole.

The hope of this dashboard is that it can be a virtual hub for cross-collaborative efforts between numerous government agencies, university scientists, and also regular citizens. Having a dashboard that utilizes the latest in machine learning to display python locations and predicted movement corridors will be a huge asset to every individual in the python removal effort, and can guide data-driven decision-making in strategy. If the python dashboard is a success, other dashboards can be created to monitor other invasive species, and potentially usher in a new method of data analysis, resource allocation, and machine learning into the field of ecology and conservation.

Sociotechnical Topic:

Providing a Data-Driven, unified Framework for Public and Private Entities to combat the threat of Invasive Species

There is a significant effort from federal, state, and local governments to get rid of the python's presence in the Everglades. However, it seems as if there are a lot of organizations doing their own approach rather than a strategic, systematic, and unified approach. This fails to account for even the interests of private conservation agencies, university researchers, and the general public. The issue of invasive species incorporates a wide variety of perspectives from financial to societal issues which further complicates the situation (Poland, 1). For example, the government might be focusing on the best use of taxpayer dollars while those in academia may be focused on using the python problem to bolster their reputation and research. All these clashing interests can impede true conservation work from taking place. Additionally, these organizations may be redoing each others' work as they lack a centralized method of communicating with each other. With a precious ecological gem like the Everglades being threatened every day by invasive species, it is important that time and resources be conserved without the threat of clashing interests. Therefore, the specific uncertainty that my STS research will provide is a data-driven, unified, and systematic approach that all researchers and those involved in the conservation effort can access and update based on their own findings. The ultimate goal of this is to bridge the public and private conservation efforts concerning the Everglades.

Failing to deal with the python problem will not just affect the Everglades but other areas dealing with invasive species. A problem with conservation is bureaucracy and red tape slowing down tangible efforts. If a collaboration framework is established that can deal with the python

problem, then there is hope for an effective way to deal with invasive species in the future. There is limited STS research that my work will build on, as my topic is relatively niche. However, I think that analyzing the organizational and cultural factors of my problem will be a good STS concept to draw from for my project. Previous research that can provide a model for this is looking at the Hurricane Katrina article we talked about in STS 4500. As shown in Figure 2, this article talked about failures within each branch of the government responsible for hurricane protection all the way from the federal to the local level (Andersen, 6). Additionally, the article revealed a deeper cultural understanding within these governmental organizations that showed a presence of cutting corners and a lack of high-quality work being produced (Andersen, 3). The article was unique in this aspect with its analysis of the organizational aspect of the governmental agencies responsible and that is something I hope to try and emulate within my own research. My hope is that by understanding the research the authors did with federal agencies responsible for hurricane protection, I can draw similar conclusions for the organizations responsible for protecting national parks from invasive species. Accomplishing this can provide a framework for identifying harmful organizational practices that can hinder meaningful work, particularly in the field of federal conservation, and also shed light on how we can improve these practices for better conservation methods.

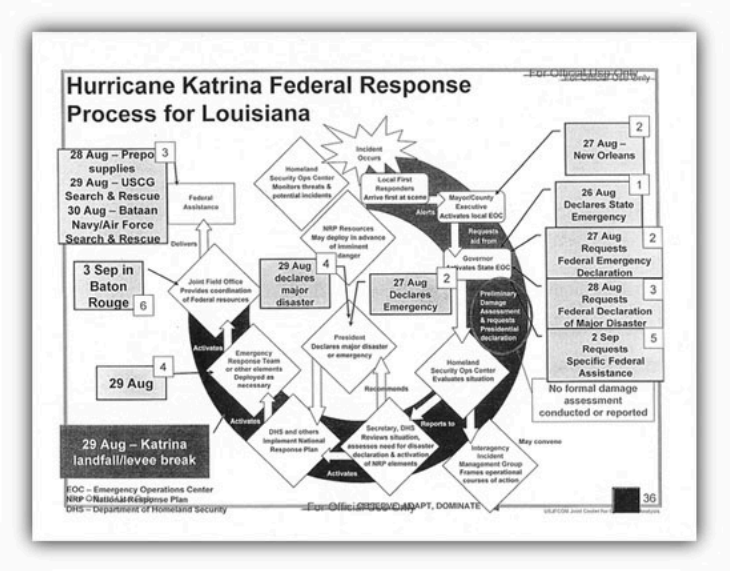

Figure 2. Bureaucracy limiting Effective Government Response (Levines 1).

Conclusion

The anticipated deliverables of my technical work would include a high-quality tableau dashboard that would list geospatial data from python movements and include high-level machine learning clustering algorithms analyzing the pythons' movements. There will be options for verified users to input python geospatial data, and explanations will be provided for non-technical users to understand the pythons movements on a broader level. It is a race against time to protect important ecosystems like the Everglades with scientists having to overcome a wide variety of technical and non-technical challenges. One of these issues is resource allocation in terms of funding and time, which can be exacerbated by the lack of cohesive work between the public and private sectors of conservation. The hope of using machine learning in this situation is far bigger than just the integration of a technical tool, but rather the providing of a framework to bridge two fields of conservation.

If my deliverables are appropriately implemented and completed successfully, this would provide a tool for researchers to better see which parts of the Everglades are most at threat from the Burmese Python. This would in turn save time and money in the struggle to save the Everglades from invasive species.

Word Count: 1745 words

References

Andersen, C.F. et al. (2006). Hurricane Katrina: One year later: What must we do next? A statement by the American Society of Civil Engineering Hurricane Katrina External Review Panel. Reston, VA: American Society of Civil Engineers.

Beaver, C. E., Meigs-Friend, G., & Hunter, M. E. (2021). Environmental DNA Surveys of Burmese Pythons in the Greater Everglades Ecosystem. Reston, Virginia: U.S. Department of the Interior, U.S. Geological Survey.

Committee on Independent Scientific Review of Everglades Restoration Progress, Water Science and Technology Board, Board on Environmental Studies and Toxicology, Division on Earth and Life Studies, National Research Council, & Ebook Central - Academic Complete (2014). Progress Toward Restoring the Everglades: The Fifth Biennial Review 2014. Washington: *National Academies Press*.

Fang, F., Tambe, M., Dilkina, B., & Plumptre, A. J. (Eds.) (2019). Artificial Intelligence and Conservation. *Cambridge University Press*.

Fort Collins Science Center, Invasive Species Science Branch (2015).: Research and Management Tools for Controlling Invasive Species. Reston, Va: U.S. Department of the Interior, U.S. Geological Survey.

Griswold, K. E., & Larsen, D. P. (2011). Developing a Survey (probability) Design to Evaluate the Presence and Distribution of Aquatic Invasive Species in Low Risk Waterbodies in the Greater Yellowstone Area. Fort Collins, Colorado: U.S. Department of the Interior, National Park Service, Natural Resource Stewardship and Science.

Hsieh, W. W. (2009). Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels. *Cambridge University Press*.

Poland, T. M. (Ed.) (2020). Invasive Species in Forests and Rangelands of the United States: A Comprehensive Science Synthesis for the United States Forest Sector. Cham: Springer.

Singh, S. (2020). Mapping Invasive Plant Species Using Machine Learning; Environmental Costs of Machine Learning Algorithms:Reducing Unnecessary Computations.

Charlottesville, VA: University of Virginia, School of Engineering and Applied Science, BS (Bachelor of Science), 2020. Retrieved from https://doi.org/10.18130/v3-fvvh-nb43

Zhang, C. (2020). Multi-Sensor System Applications in the Everglades Ecosystems. Boca Raton: *CRC Press LLC.*