

COMP-551: Applied Machine Learning

Project 4 : Improving Conservation of Animal Species

Due on November 29 (short oral presentation) and December 9 (written report)

Background:

There are many opportunities to apply machine learning techniques to tackle challenging data analysis problems pertaining to the environment and conservation. As an example, here is a recent Kaggle competition on this theme: <https://www.kaggle.com/c/the-nature-conservancy-fisheries-monitoring>

The goal of this project is to investigate animal tracking data from the MoveBank database: <https://www.movebank.org>. Begin by identifying an interesting dataset and prediction question that can be tackled using machine learning methods, and solve the problem using appropriate machine learning algorithms and methodology. You must use at least one dataset from MoveBank. Additionally, you can incorporate data from other sources (e.g. similar animal tracking data, weather, geography, etc.)

The choice of prediction task is open. Try to pick a prediction question that is relevant and important to improving conservation and our understanding of animal behavior. Remember to design a prediction task that is well suited to your choice of dataset; and vice versa, pick the right data for tackling your prediction question.

There are no mandatory algorithms to implement for this project. You can implement your own algorithms, or use any existing toolbox or software, as long as you reference them appropriately in your report. Do not hesitate to explore new models, algorithms and architectures that are appropriate for this type of data.

The final project should be completed in groups of three. As usual, you must work with people with who you have not worked with during previous projects.

Requirements:

There are 3 components to submit:

1. Short oral presentation (3min / team) to be presented in class on November 29. One or all team members can speak. Slides can be used (pdf format only). Presentations will be timed precisely, so practice accordingly. The presentation should clearly identify the target prediction question, describe the data used, define the proposed algorithms, explain the proposed evaluation methodology, and if possible, give preliminary results. Evaluation criteria: Importance of prediction question, pertinence of methods, discussion of results or expected results, quality of delivery, organization. Weight: 10% of final project grade.
2. Written report (1 / team). The report should have at most 8 pages. It should follow a standard conference format (see instructions from previous projects). The report should clearly present the target prediction task, the data used, a clear and well-motivated methodology for analyzing the target question with the selected data, empirical results of the analysis and a discussion. When appropriate, related work (e.g. attempts to tackle a similar prediction question, or analyze similar data) should be discussed and referenced. Weight: 90% of final project grade.
3. Code (1 / team). Code developed for this project should be uploaded as supplementary material.

Submission instructions:

Slides for the oral presentations should be emailed to the instructor by 12pm on November 29. The final report & code should be submitted on CMT, as with previous projects.