Project I

Modeling urban coastal flood severity from crowd-sourced flood reports using Poisson regression and Random Forest

Part I: Paper Replication

1. Read this paper, discuss it with your group and replicate this study on cloud (i.e., AWS-EC2).
2. Report any challenges with replicating this work, as well as the findings, ease of replication, what needed to be changed and any other information that was missing to replicate this paper. All script and data used for this paper are shared through HydroShare Resources, and are also available on GitHub (Fork the paper repository). A collection resource was created to aggregate all the resources related to street flood severity modeling in Norfolk, Virginia USA (raw and pre-processed data, scripts used to perform the pre-processing, scripts used to train data-driven algorithms, and results from the models).
3. The author of this paper shared all the script used from pre-processing to generate the model output through HydroShare using different types of resources within HydroShare to make his work reproducible and transparent. From your experience in replicating this paper, state the pros and cons from using HydroShare.
4. In your opinion, what additional functionality should HydroShare implement to support the use of HydroShare to make research more reproducible and transparent?

Part II: Paper Reproducibility

1. With your group, discuss ideas of how you can enhance reproducibility of this paper and implement one of these ideas. As an example, reproducibility can be achieved by generating a workflow to automate previous work, packaging the workflow with all its dependencies using container technology. Then the packages can be run on JetsStream or even Amazon Web Services (AWS-EC2).
2. Generate a workflow diagram of your reproducibility plan for this paper (example: workflow diagram)?

Part III: Workshop Outcomes

1. From what you have learned so far, write a reproducibility plan you think other scientists can follow to achieve reproducible research. Include any tools that are applicable to your research, and how they could be used to enhance reproducibility in your own area of study. (max. 2 pages)
2. Create a presentation to summarizing your work on this project. This should include:
    a. The problem statement.
    b. How you were able to replicate this paper
        i. Any challenges and how they were addressed
    c. Outcome of the replication:
        I. pros and cons of using HydroShare
        II. Any functionality that could address issues with HydroShare
    d. How you were able to reproduce this paper (what steps were taken, tools were used, challenges and how you addressed these challenges)

e. The workflow diagram

f. Summarize your outcomes from this two weeks' workshop, including any software or tools you could use in your own research.