GGE6404: Online Spatial Data Handling

INSTRUCTIONS

Explore data and create predictions

In this assignment, you will create predictions and create a visualization of the prediction in Jupyter Notebooks.

This final assignment report should contain the 8 phases of KDD as discussed in Module 6, Lecture 2 and access data that is available online.

Thus, sections of the report should include:
a) An introduction to the project and question being answered (e.g.: what areas are susceptible to flooding)
b) Details on the dataset on which the discovery is being performed (dependent variable) (e.g.: source, data, projection, format and what processing might have been done on it)
c) Data preprocessing details on the independent variables that are included, including source and why it was selected)
d) The task that has been selected and why, e.g: are you performing classification or regression
e) The selection of the algorithm and why, e.g.: CART, random forest, etc.
f) Running the algorithm and any hyper-parameterization
g) Evaluation of the results – include accuracy, ROC-AUC, and at least one other measure from the confusion matrix and explain what they represent
h) Finally, run the model on-some untrained data and present the results in a map; the dynamic map should include the results layer as well as layers of the independent variables used in the modelling which can be turned on/off

You will submit a PDF result of your work as well as the notebook itself (ipynb file).

You must complete one of the following two options for this assignment:

Option 1: Instructor supplied topic: Flood Susceptibility predictions

In New Brunswick, there is a long history of flood events along the Saint John River, especially in the lower section of the river, between Fredericton and St. John. In this assignment, you will use a historic event of flooding to predict the probability of flooding anywhere in your study region.

You may want to look at a watershed area, like Swan Creek (pictured below) and predict the flood susceptibility anywhere in the watershed. You can use a historic flood event from NRCan

Floods in Canada – Cartographic Product Collection
(https://open.canada.ca/data/en/dataset/08b810c2-7c81-40f1-adb1-c32c8a2c9f50) as the
dependent variable.
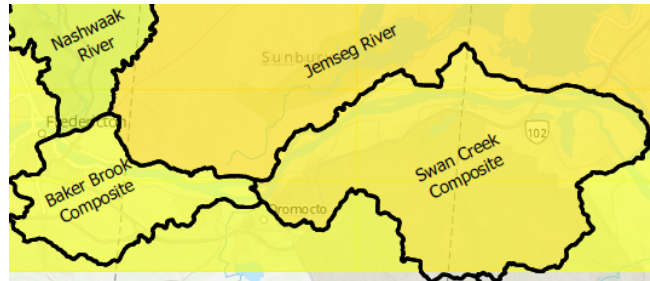
Please note that the polygon boundaries and names in the figure above are from GeoNB Data
Catalog: http://www.snb.ca/geonb1/e/DC/catalogue-E.asp

Options for independent variables include (*you should include at least 3):
- Digital elevation model (DEM)
  - Derivatives of DEM
  - Slope or roughness
- Land cover; https://open.canada.ca/data/en/dataset/4e615eae-b90c-420b-adee-2ca35896caf6
- Forested areas or forest soils from GeoNB; http://www.snb.ca/geonb1/e/DC/catalogue-E.asp
- Historic weather data from GeoMET API; https://eccc-msc.github.io/open-data/msc-geomet/readme_en/
  - Precipitation or snow variables could be considered.

Option 2: Student supplied topic:

I welcome you to select your own topic for this assignment. You should consider at least 3
independent variables in your analysis, but the topic it is up to you, based on your own interests.
The topic should be looking to solve a spatial problem and include access to at least 3 different
datasets.

If you select this option, you must contact me in advance to discuss if a particular topic is suitable.

EXPECTATIONS

You will need to submit:
  a) Jupyter Notebook (ipynb) file (for me to view/run) and
  b) A PDF or HTML of the Jupyter notebook which contains the output as run by you

The Notebook document should read like a report where you include Markdown to supplement the
code.

At the top of the submission, you should include:
- The assignment name and description
- Your information (name, student number)
- The date

The submission should include runnable code cells and a brief description, in Markdown, of:
- The description and objective of the assignment
- The steps you performed, libraries and functions used and why
- Links and references to all data used

Where appropriate, you should include charts, graphs, maps, interactive maps and formatted text to summarize and describe the results.

I intend to run the submitted Notebook, so please ensure all libraries and necessary files are included/referenced.

The assignment will be graded based on the rubric below.

Assignments - Grading Rubric					
Criteria	Level 1	Level 2	Level 3	Level 4	Level 5
Code Functionality (4%)	None of the code runs successfully	The code partially runs successfully	The code produces incorrect results	The code generally produces the right results	The code works well and meets all specifications
	0 point	1 point	2 points	3 points	4 points
Readability and Compactness (2%)	The code is poorly organized, lots of unnecessary lines and is incomplete	The code is poorly organized and difficult to read, lots of unnecessary lines of code	The code is only readable by someone who knows what it should do, unnecessarily long	Code is fairly easy to follow, but not well organized, a little verbose	Exceptionally well organized, easy to follow, concise
	0 point	0.5 point	1 points	1.5 points	2 points
Overall Assignment Organization and Documentation (4%)	No comments in code and no use of Markdown	Only comments embedded in the code and little to no use of Markdown	Limited use of Markdown to explain the assignment objectives and support the work completed	Assignment objective is included, project is described, but there are missing method/tools used, figures, references	Assignment objective is well described, including all steps. The code is supplemented with description of methods,

	figures, references to resources used, etc.
0 point	1 point	2 points	3 points	4 points

D2L DROPBOX SUBMISSION INSTRUCTIONS

1. In the top-navigation bar on the course screen, select 'Assessments' and then 'Assignments'.
2. Select the assignment title and follow the instructions to upload your document.