Introduction and Background

Satellite remote sensing of the earth is an effective way to gather data over a large surface area. This data can be used for various assessments, such as determining the vegetative health of a region, the richness of the soil, or the composition of the landscape. Land classification models discern which areas within a scene belong to which land cover type; such as, coniferous vegetation, deciduous vegetation, wetlands, water bodies, urban areas, agriculture, grasslands, or desert. This can then be used for land development, natural resource management, and impact assessment studies [1]. These land classification algorithms help to turn satellite imagery into a tangible product which can then be compared to other data. As sensors become more spectrally and spatially refined, more accurate and detailed accounts for land cover are possible. Geographers are starting to take a 'hands off' approach to analyses and trust in computing power to derive their products. Machine learning and artificial intelligence is emerging as a large player in perfecting models for geography analyses. This study will compare probabilistic, support vector machine, and neural network models. As there are many avenues which have been explored, and many different models to compare to, this project will start with a broad scope and become more focused with time.

About Me

As a student in the computer science and geography field, it interests me how the two disciplines can come together. Some of my favorite computer science and software engineering classes were Data Mining and Introduction to Artificial Intelligence. For this reason, I would like to combine various machine learning classification algorithms with remote sensing data. My background as a Junior Remote Sensing Developer with Natural Resources Canada has provided me with the basic understanding on how to approach this project and some connections to obtain the help I may need.

Andrea Nesdoly (V00767633) - BSc. Major Combined Computer Science and Geography

Methods and Milestones

This section describes the various methods and milestones to perform the land classification algorithm comparison project. It explains the data source and contains a broad overview on the goals of the project, it does not go into too much detail to allow the project some fluidity.

Data

Considerations when procuring data for this project will be temporal and spatial resolution of the data. Data from the summer months will have a larger variety of vegetation than data from the winter months and finer spatial resolution will allow for more detailed and accurate classification because there will be less pixel mixing. Multiple data sources may be considered so long as they are open and available.

The study area is the Saanich Peninsula which has a complex and heterogeneous landscape. It contains various agriculture, urban, and water based landscapes which will lead to a more difficult and interesting classification. The data displayed below is available through the U.S Geological Survey, and was gathered using the Sentinel-2A platform [2,3]. The tile gathered using the Sentinel-2 was acquired on November 11, 2018 on a descending orbit. The spatial resolution of this product is 10, 20, or 60 meters depending on which spectral bands are being considered. The latitude and longitude of the centre of the tile is 48°15’27.01”N and 123°36’29.39”W respectively. As the study area will consist of only the Saanich Peninsula this tile will be trimmed before processing.

Figure 1: Sentintel-2 remote tile of Southern Vancouver Island.

Methods

Retrieving data and processing the tile using Geomatica, Google Earth Engine and/or ENVI software will be the first step before creating programs for land classification. Once the tile has been processed the software will be used for a land classification which will be considered the base case or truth value. Each time the software performs a land classification the output could change slightly; therefore, a few simulations may be performed and a range or average pixel value will be obtained for the final comparison.

Next, a probabilistic land classification model will be programed which will be compared pixel by pixel against the software output. Then statistical analysis will be performed to derive a r-square and root mean square error between the two. Following this a neural network and a support-vector machine (SVM) land classification model will be performed. The output for these algorithms will be compared to the base case in a similar way. Idealistically, variable training sample sizes will be considered. Along with statistical analysis, accuracy assessments, time efficiency, and space efficiency will be compared between algorithms. There is hope that this project may lead to a simulation of other known models, like a self-trained semisupervised SVM [4], a pixel versus object based classification method [5], or the use of multiple data sources.

Jupyter notebooks and Github will be used to process and share the project; and Google Earth Engine will be considered for data acquisition and available cloud computing technologies.

Project Gantt Chart

Below is a gantt chart of the various project milestones displaying when certain tasks occur and when they should be completed by. As a the date for the final deliverables and presentation have not yet been set this may change slightly.

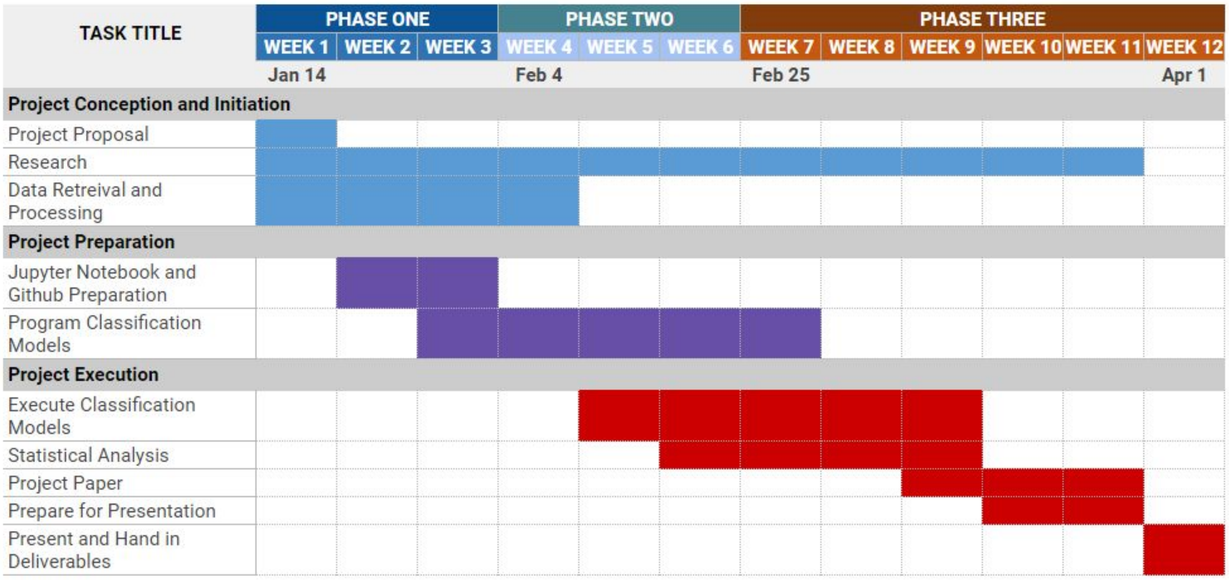

Figure 2: Gantt Chart

References

Gall, M., Borden, K. and Cutter, S., 2009: *When Do Losses Count? Six Fallacies of Natural Hazards Loss Data*, American Meteorological Society. DOI:10.1175/2008BAMS2721.1

[1] Noi, P.T. and Kappa, M., 2018: *Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery*, Sensors. DOI: 10.3390/s18010018

[2] Data Set: https://earthexplorer.usgs.gov/

[3] MetaData: https://earthexplorer.usgs.gov/metadata/10880/6052370/

[4] Lui, Y, Zang, B., Wang, L., and Wang, N., 2013: *A self-trained semisupervised SVM approach to the remote sensing land cover classification*, Computers and Geosciences. http://dx.doi.org/10.1016/j.cageo/2013.03.024

[5] Duro, D.C., Franklin, S.E., and Dube, M.G., 2011: *A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery*, Remote Sensing of Environment. DOI: 10.1016/j.rse.2011.11.020