



landslide-ml-tutorial / README.md



indraneelk Update README.md

1 hour ago



47 lines (28 loc) · 3.6 KB

Preview

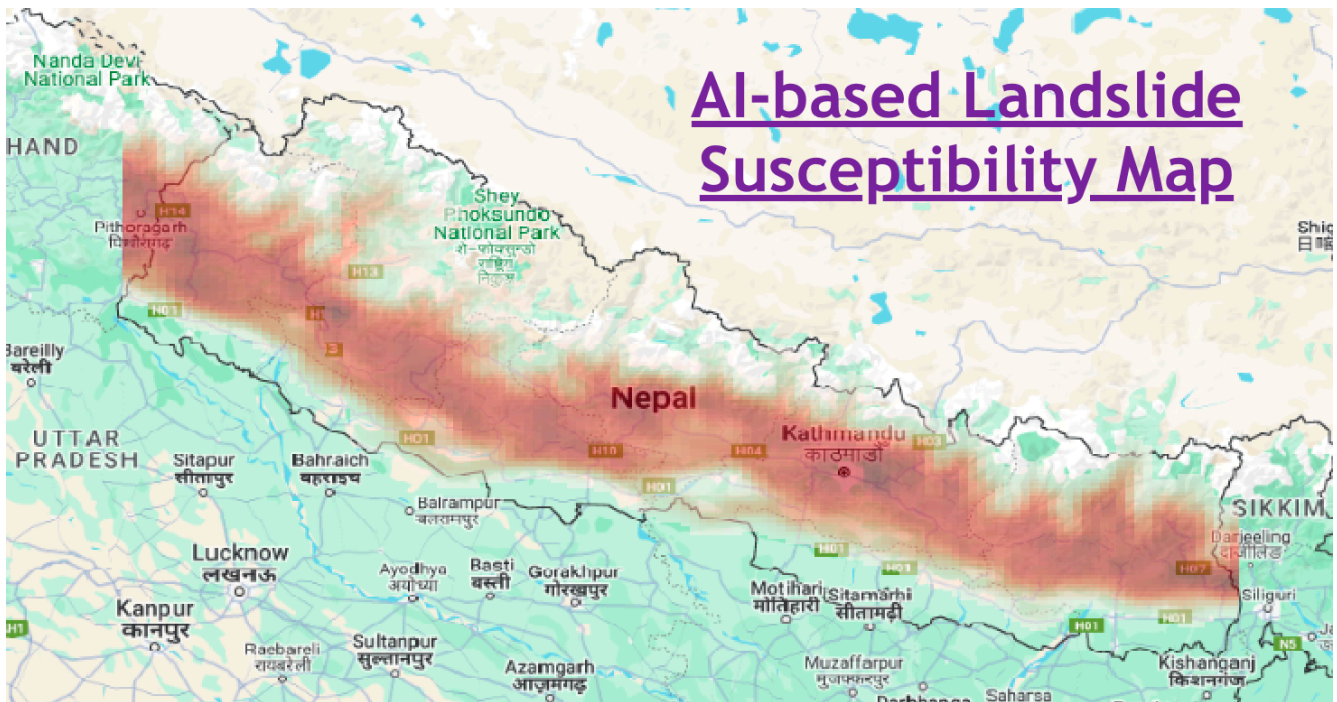
Code

Blame

Raw



Machine Learning Tutorial: Landslide Edition



Introduction

This repository is part of the Tuesday May 21 hands-on AI/ML Basics Tutorial. The goal is to learn the basics of machine learning through the process of building simple landslide models.

Objectives

1. Build a logistic regression model and a random forest model for landslides in R.
2. Learn about training/testing, hyperparameters, and overfitting.
3. Learn how to add new datasets to the existing dataset to improve the models.
4. Build your own landslide susceptibility map using the random forest model.

Prerequisites

Follow the steps from [system-setup](#) to ensure you have all the necessary modules.

1. Set up Github environment for code management (highly recommended if you will be working on coding projects, but not necessary).
2. Set up QGIS for spatial data viewing.
3. Set up R programming language.
4. Set up python programming language (optional).
5. Download the code to your local computer from either the [Github](#) (if you already have a GitHub account), or from the [Google Drive](#).

Also think about what you could do with this framework to pursue your own questions - a wildfire model, a pollution model,

Logistic Regression Model for Landslides in Ecuador

We first learn about the basics of R and logistic regression as we build a model for landslides in Ecuador. Open the [landslide_logistic_regression.Rmd](#) in RStudio and run each code block.

Logistic regression is a simple model that, given certain inputs, provides the probability of landslide as an output. It estimates the probability by looking at patterns in the data. It is like drawing a line through data (regression), except the y variable is transformed to be a probability (between 0 and 1).

There are certain exercises within the code. So make sure to do them.

Random Forest Model for Landslides in Nepal

We then build a random forest model for landslides in Ecuador. Open the [landslide_randomforest.Rmd](#) in RStudio and run each code block.

A random forest model is a collection of decision trees. If you provide landslide input data and an indicator of landslide (yes/no), then a computer is capable of generating a decision tree to fit that data (e.g. if slope > 45, then landslide likely. If elevation > 1000 m then landslide likely...). A random forest is a collection of 100s of such decision trees. Given input data, the random forest asks each of its decision trees to give its response, and then returns the popular vote (landslide yes/no).

Random Forests are one of the most versatile and robust machine learning models out there, and quite simple to build compared to the intense neural network models.

How was landslide data generated for Nepal?

The landslide data is simply generated from terrain characteristics. The [MERIT](#) Digital Elevation Model (DEM) was used to obtain elevation, slope and curvature data. And it turns out, this data is sufficient to make a decent landslide prediction model.

The actual data generation was done with [data_prep.Rmd](#). This dataset collects landslides from the Global Landslide Catalog, then generates non-landslide points, generates terrain data, and compiles the dataset. You can alter this R notebook to get suitable data for yourself.

