Wildfire Detection

Mentors: Hunter McCoy & Nandita Radhakrishnan

Team: Vaibhav Mohata, Pratyush Das, Shani Lyubomirsky, Neel Datta, Peter F. Grinde-Hollevik, Sifath Mannan, Aashritha Srirambatla

University of California Berkeley Undergraduate Data Science Lab

Background

Wildfires are a growing natural hazard in most regions of the United States, posing a threat to life and property, particularly where native ecosystems meet developed areas. (US Geological Survey, 2006). The societal consequences are clear: Just in the United States alone, over 13 million acres of land burned, $3.2 trillion were spent in wildfire suppression costs, and over 14.700 structures were lost in the 2020 wildfire season. (National Large Incident Year-to-Date Report, 2020).

While the current research is good at generating models for current or immediately-close fires, there is still a gap in the use of data in considering the effects of climate change on the intensity of wildfires.

We look to close that gap with our research through the development of our prediction engine.

Goals

1. Identify the instantaneous and future likelihoods of fires in a certain location based on regional weather and vegetation patterns.
2. Achieve a higher prediction accuracy using a temporal data-based approach in comparison to current measurements of wildfire prediction.
3. Train the model from scratch to achieve up to 90% test accuracy.

Data

The data we used to train our models were obtained from the MODIS LSAT, Vegetation, and Fire 8-Day V3 datasets. MODIS is the Moderate Resolution Imaging Spectroradiometer which captures images of the Earth with 36 discrete spectral bands. MODIS data has three spatial resolutions: 250m, 500m, and 1000m.

LSAT data was obtained during daytime and nighttime with normalized vegetation index data (NDVI).
The MODIS Vegetation data contains minimized canopy/ soil variations and improved sensitivity to dense vegetation.
The Fire dataset contains recorded fires in 1-km pixels that were burning at the time of overpass.

For the purpose of our research, we focused on image datasets obtained from Eastern Australia, a portion of the globe frequently affected by fire, over a period from May 2018 to June 2020. The data contained the latitude and longitude of specific coordinates and the spectral information for each pixel in the image dataset, packaged into an HDF file.

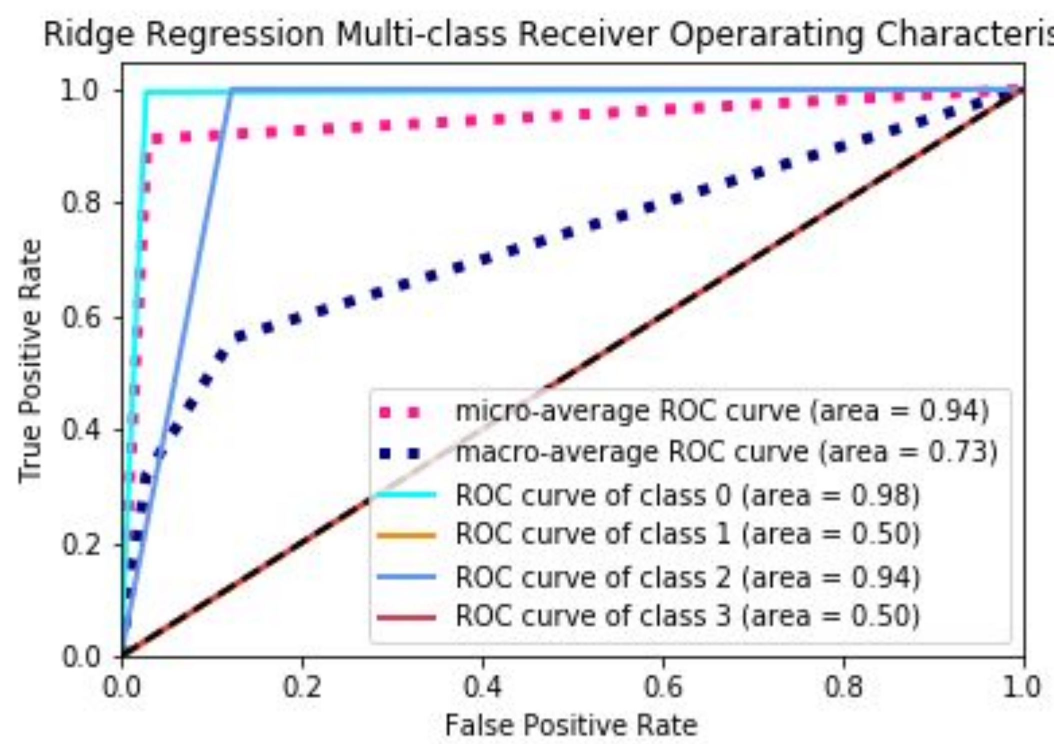
Data Cleaning and Aggregation

Data processing was performed using the Pyspark library. The raw HDF file required a specialized reader, so we mapped them to csv files and then applied a reverse transform to extract a mapping from pixels to geographic coordinates. We then performed an average pooling to reduce the scale of the dataset, joined based on temporal and spatial locality, and removed samples with incomplete or low accuracy data in the process. The cleaned aggregation was split into training and validation sets and then boosted to ensure equal numbers of all samples.

Models and Evalution

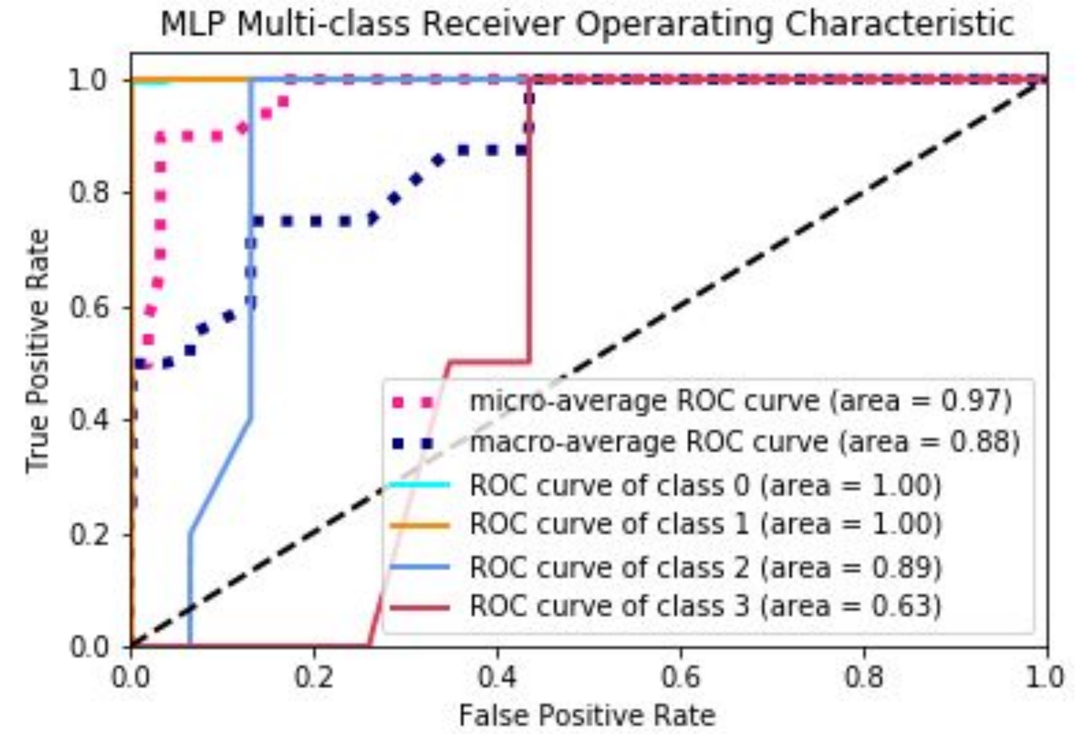
Ridge Regressor

Our first model was a Ridge Regressor as we assumed the data was linearly separable and still possessed some small error. This model had a test accuracy of 91.5% and a macro AUC of .73.



Multi-Layer Perceptron

The final model we chose was a 4 layer categorical MLP trained with Adagrad. To determine our hyperparameters, we partitioned the data into training, test, and validation sets, and performed cross-fold validation to locate the ideal hyperparameters.



Summary

The final model achieves an average of ~91.33% accuracy and ~88% macro AUC. Due to limited time and resources, it is difficult to train a model from scratch to reach optimal performance. Even with a complex architecture, the model had difficulty surpassing the threshold in its predictions. With more time and budget for training, we are confident we can increase AUC to a near-optimal level.
From this project, we learned how to use Amazon AWS, MODIS, and Pyspark to build a large scale model for wildfire detection.

Next Steps

The next step for the project would be to rebuild the dataset aligned to a 3d grid of latitude/longitude/date, so that we could use a convolutional recurrent network. In addition, we would like to expand the reach of our data to encompass the entire world instead of just eastern Australia.

References

U.S. Geological Survey, U. S. D. of the I. (2006). Wildfire Hazards—A National Threat

National Interagency Fire Center. (2020). National Large Incident Year-to-Date Report.

Walsh, C. (2020, January 31). The long-term effects of wildfires. Harvard Gazette.

Y.O. Sayad, H. Mousannif, H. Al Moatassime, Predictive modeling of wildfires:A new dataset and machine learning approach, Fire Safety Journal (2019), doi:
Coen, J. Douglas, C. (2020) Computational Modeling of Large Wildfires: A Roadmap

P. Cortez and A. Morais, "A Data Mining Approach to Predict wildfires using Meteorological Data," Proc. 13th Port. Conf. Artif. Intell., pp. 512–523, 2007
O. Rios, E. Pastor, M.G. Valero, E. Planas (January 2016) "Short-term fire front spread prediction using inverse modelling and airborne infrared images"