MACHINE LEARNING & TEMPERATURE CHANGES

Dalit Hendel

DATA 200 Final Project

December 22, 2021

GOALS:

- Compare the accuracy of a Random Forest and a Neural Net model both trained on WorldClim 2.0 (historic and future) data to identify land surface areas that have had a temperature change of more than 1.5°C.
- Gain insights into predicting possible temperature and climate trends & use this gained knowledge to influence actionable policy change and intervention on the government level.

MODELS:

Neural Net

Random Forest

INTRODUCTION:

- Understanding and mitigating the effects of climate change is the most pressing issue of the 21st century. Global-scale concerns that should have been addressed decades ago linger and are swept under the rug for future generations to tackle. Lack of data is not the cause of the lack of climate change mitigating action. Using this previously collected data for meaningful analysis is paramount to better understanding past and future climate trends and preventing/mitigating their adverse effects.

DATA:

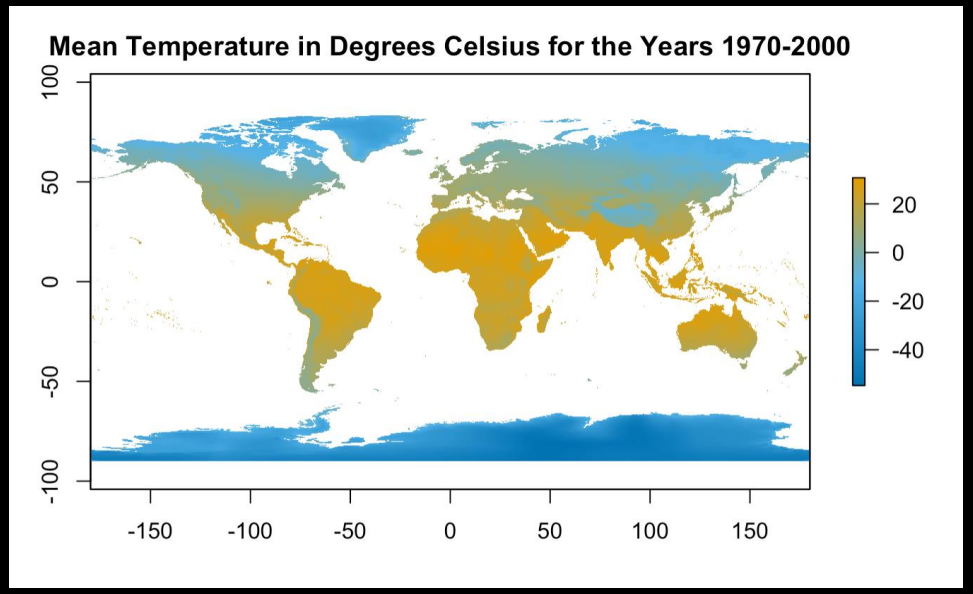
Historical climate change (1970 – 2000)

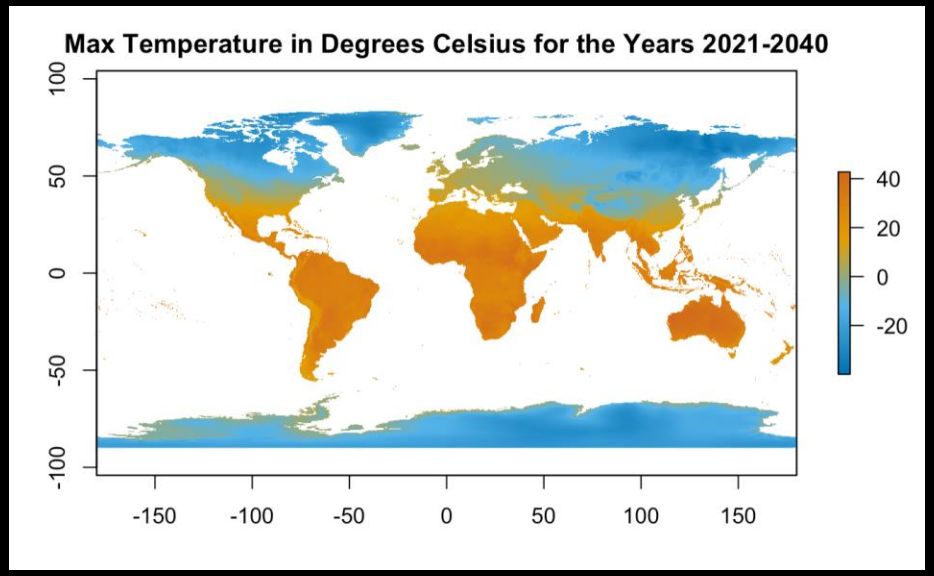
- Longitude, Latitude, Average temperature (°C), Precipitation (mm), Elevation
- 10 minutes (~340 km2)

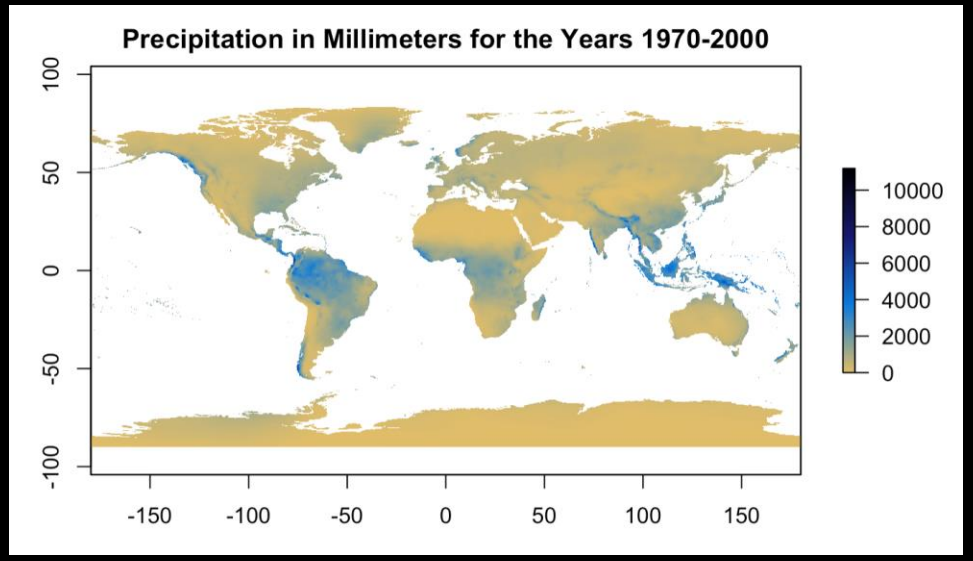
Future climate data (2021 – 2040)

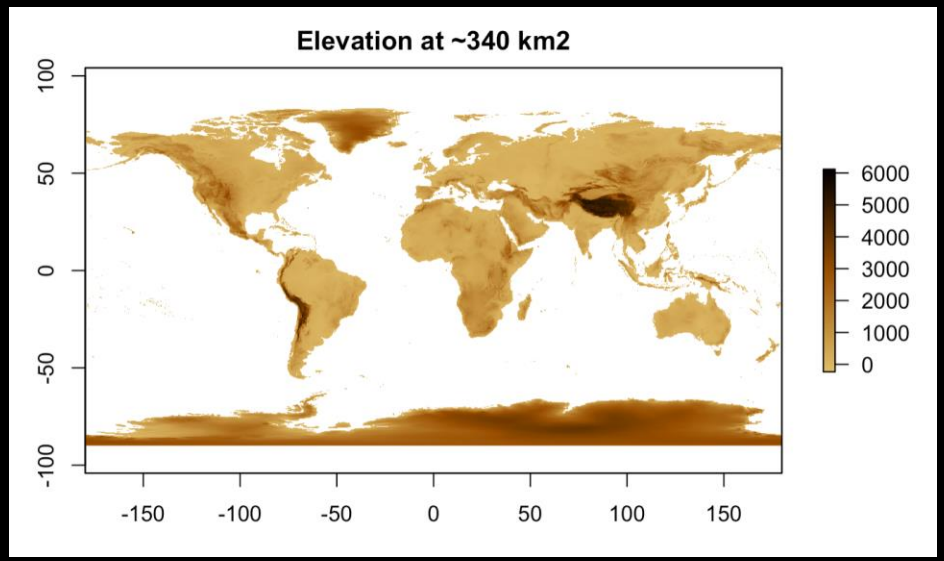
- Longitude, Latitude, Max temperature predictions (°C)
- 10 minutes (~340 km2)
- Estimates from RCP2.6 emission and concentration pathway modeled to show potential climactic outcomes of successful climate mitigation for which the global average temperature does not rise more than 2°C (van Vuuren et al. 2011)

VISUALIZING THE VARIABLES:







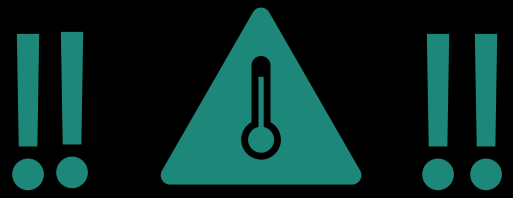


MODEL DETAILS:

X = mean temperature (1970-2000), precipitation (1970-2000), elevation

y = Δ in temperature

- Δ = max temperature (2021-2040) - mean temperature (1970-2000)
- 0 if Δ < 1.5℃ & 1 if Δ > 1.5℃
- Repeated cross-validation including 10 folds and 5 repeats



DATASET & LIMITATIONS:

- The data frequency for both data sets is at the spatial resolution of 10 minutes (~340 km2) and was downloaded as GeoTiff (.tif) files
- Each .tif file contained one layer of information which was extracted in RStudio using the package raster.
- After extracting longitudinal and latitudinal information, each dataset contained 2,332,800 rows. After dropping empty rows (longitude and latitude associated with non-land surface areas of the planet) the data frame contained 80,5693 rows
- HAD TO BE CUT DOWN DUE TO COMPUTATIONAL LIMITATIONS
- The final data frame size used in the model was 5,000 rows x 6 columns (the largest slice that would run on my computer)
- Left the dataset with 4,868 target values where the temperature change > 1.5℃ and 132 target values where the temperature change < 1.5℃. (Imbalanced)

EMPIRICAL ANALYSIS & FUTURE WORK

- Most errors were type 1 (false positives)
- RF model performed better in Accuracy and Specificity
- Imbalanced target variable is reflected in the low Specificity scores
- Better distribution of target variables in training data needed

  - oversample minority class & undersample majority class
- Access to a high-speed computer to run full data

Neural Net
Accuracy: 97%
Sensitivity: 99%
Specificity: 5%

Random Forest
Accuracy: 98%
Sensitivity: 99%
Specificity: 29%

IN CONCLUSION...

GOAL: use available environmental data to test several machine learning models and their abilities to predict which areas will be vulnerable to the greatest temperature changes.

The necessity of gaining these insights is paramount to instigating actionable policy change and interventions to the regions that will be the most at risk

This model is run on some of the most optimistic climactic outcomes for which the global average temperature does not rise more than 2°C. These models' predictive relevancy to the real world are dependent on the global average temperature not increasing by more than 2°C. For this to be possible immediate action should be taken to reduce greenhouse gas emissions on a global and local scale.

THANK YOU!

REFERENCES:

- The Climate Reality Project. (2021, January 25). Why is 1.5 degrees the danger line for global warming? Climate Reality. Retrieved December 15, 2021, from https://www.climaterealityproject.org/blog/why-15-degrees-danger-lineglobal-warming
- Gleckler, Durack, P. J., Stouffer, R. J., Johnson, G. C., & Forest, C. E. (2016). Industrial-era lobal ocean heat uptake doubles in recent decades. Nature Climate Change, 6(4), 394 398.https://doi.org/10.1038/nclimate2915
- Held, & Soden, B. J. (2006). Robust Responses of the Hydrological Cycle to Global Warming. Journal of Climate, 19(21), 5686–5699. https://doi.org/10.1175/JCLI3990.1
- Keeling, Bacastow, R. B., Bainbridge, A. E., Ekdahl Jr, C. A., Guenther, P. R., Waterman, L. S., & Chin, J. F. S. (1976). Atmospheric carbon dioxide variations at Mauna Loa Observatory, Hawaii. Tellus, 28(6), 538–551. https://doi.org/10.3402/tellusa.v28i6.11322
- Manabe, S., & Wetherald, R. T. (1967). Thermal Equilibrium of the Atmosphere with a Given Distribution of Relative Humidity, Journal of Atmospheric Sciences, 24(3), 241-259. https://doi.org/10.1175/1520-0469(1967)024<0241:TEOTAW>2.0.CO;2
- Neukom, Barboza, L. A., Erb, M. P., Shi, F., Emile-Geay, J., Evans, M. N., Franke, J., Kaufman, D.S., Lucke, L., Rehfeld, K., Schurer, A., Zhu, F., Bronnimann, S., Hakim, G. J., Henley, B. J., Ljungqvist, F. C., McKay, N., Valler, V., & von Gunten, L. (2019). Consistent multidecadalvariability in global temperature reconstructions and simulations over the Common Era. Nature Geoscience, 12(8), 643–649. https://doi.org/10.1038/s41561-019 0400-0
- Unfccc.int. (n.d.). Retrieved December 15, 2021, from https://unfccc.int/process-and meetings/the-paris-agreement/the-paris-agreement
- van Vuuren, D.P., Stehfest, E., den Elzen, M.G.J. et al. RCP2.6: exploring the possibility to keep lobal mean temperature increase below 2°C. Climatic Change 109, 95 (2011). Https://doi.org/10.1007/s10584-011-0152-3
- WorldClim (2020). Historical climate data, 1970-2000. Version 2.1. https://www.worldclim.org/data/worldclim21.html Web. 12 Dec 2021.
- WorldClim (2020). Future climate data, 1970-2000. Version 2.1. https://www.worldclim.org/data/cmip6/cmip6climate.html Web. 12 Dec 2021.
- Statistics were done using R 3.5.0 (R Core Team, 2018), the dplyr (v1.0.7; Wickham, 2021), the raters (v2.0.1; Giardiello, 2014) packages, the nnet (v7.3; Ripley and Venables, 2021) packages, and the randomForest (v4.6; Liaw and Wiener, 2018) packages.