

# Predicting Climate Change Case Study

**DS 4002 – Fall 2023**

**Due: TBD**

**Submission Format:** link to [GitHub Repository](#) on Canvas

## Individual Assignment

**General Description:** Submit to Canvas a link to your GitHub Repository

Preparatory Assignments – class discussion and assigned reading

**Why am I doing this?** The goal of the assignment is for you to analyze time series data with different models, and you will learn about how to test and evaluate them to find the best model for your purpose. You will be able to use the data science skills you have learned at every step of the process, from cleaning data to creating graphs to validating results. By completing this study, you will have applied your skills to a real-world issue like climate change and gain experience in how data science is used to solve serious problems.

- Course Learning Objective: building and evaluating models
- Course Learning Objective: summarize findings to explain project to an audience

**What am I going to do?** You will begin by reading the hook document to understand the assignment and the deliverables. Then you will look at the provided datasets and select a month and region. You have the choice to choose whichever place interests you the most! Just remember that your interpretation of the analysis will change depending on the climate of the region you choose. Perform exploratory analysis and include any produced graphs or figures in the GitHub repo. Then, test and validate different models. Select an appropriate modelling approach once you have evaluated the best one. Each model will have pros and cons, so you will have to decide what is more important to have. Justify the decisions you make in the comments. Using the selected model, predict the temperature anomalies for the years 2020, 2021, and 2022 and compare it to the observed anomalies to determine accuracy of the model. You will upload all relevant code with comments to the repo. Write a paper summarizing the methodology and results of the project and upload it to the repo.

All of this will be submitted electronically via a link to a GitHub repository.

### Tips for success:

- Heavily comment your code as you go. This will make you think about your decisions and help you explain your methodology in your paper.
- Don't overthink which model is best. It does not need to be 100% accurate; just understand its limitations.
- Have fun! You have a lot of freedom to select the dataset and model, so pick what is most interesting to you!

**How will I know I have Succeeded?** You will meet expectations on Climate Change Case Study when you follow the criteria in the rubric below.

Spec Category	Spec Details
Formatting	<ul style="list-style-type: none"> <li>• A GitHub repository (submitted via a link on Canvas) <ul style="list-style-type: none"> <li>○ README.md file</li> <li>○ LICENSE.md file</li> <li>○ SRC folder</li> <li>○ Data folder</li> <li>○ Summarizations and Conclusions document – 2-page pdf</li> </ul> </li> </ul>
README.md	<ul style="list-style-type: none"> <li>• Goal: This file will tell people who come to your repository what its contents are and how to navigate it.</li> <li>• Use markdown headers to divide sections.</li> <li>• Make an H2 section listing the contents of the repo.</li> <li>• SRC section <ul style="list-style-type: none"> <li>○ Make an H3 section for Installing/Building the code.</li> <li>○ Make an H3 section for Usage of the code.</li> </ul> </li> <li>• Data section <ul style="list-style-type: none"> <li>○ Include all data files used.</li> <li>○ Data Dictionary (use markdown table formatting)</li> <li>○ Any exploratory graphs created</li> </ul> </li> <li>• References section <ul style="list-style-type: none"> <li>○ All references should be listed at the end of the README.md file (Use IEEE Documentation style)</li> <li>○ Include any acknowledgements</li> </ul> </li> </ul>
LICENSE.md	<ul style="list-style-type: none"> <li>• Goal: This file explains to a visitor the terms under which they may use and cite your repository.</li> <li>• Usually, the MIT license is appropriate.</li> </ul>
SRC Folder	<ul style="list-style-type: none"> <li>• Goal: This folder has all the code for the project.</li> <li>• Include all code produced with comments. <ul style="list-style-type: none"> <li>○ Code for data exploratory analysis</li> <li>○ Code for model</li> <li>○ Any additional code</li> </ul> </li> </ul>
Data Folder	<ul style="list-style-type: none"> <li>• Goal: This folder has all the data for the project.</li> <li>• Place the code directly into GitHub or include the link to the data.</li> <li>• Data Dictionary</li> <li>• Exploratory graphs and explanations of them</li> </ul>
Summarization and Conclusions Document	<ul style="list-style-type: none"> <li>• 2-page PDF</li> <li>• Explain the region and month chosen and any variations specific to those choices.</li> <li>• Describe the tested models and justify the selected model.</li> <li>• Include figures and conclusions you can draw from them.</li> </ul>

Acknowledgements: Special thanks to Jess Taggart from UVA CTE for coaching on making this rubric. Special thanks to Peter Alonzi and Harsh Anand for their instruction this semester. This structure is pulled direction from [Streifer & Palmer \(2020\)](#).