- **Title**: Summarizes the main idea of your project.
  - Ice Melting Predictions using LSTMs and 2D CNNs
- **Who**: Names and logins of all your group members.
  - Kamya Raman - kraman4
  - Jinho Lee - jlee812
  - Duru Huseyni - zhuseyni
- **Introduction**: What problem are you trying to solve and why?
  - If you are implementing an existing paper, describe the paper's objectives and why you chose this paper.
    - Objective: Predict the future of sea ice melting/surface area decrease to better understand climate emergency
    - Motivation: Sea ice has been melting at unprecedented rates that should be better understood to create targeted solutions.
  - What kind of problem is this? Classification? Regression? Structured prediction? Reinforcement Learning? Unsupervised Learning? Etc.
    - Structured prediction
- **Related Work**: Are you aware of any, or is there any prior work that you drew on to do your project?
  - Please read and briefly summarize (no more than one paragraph) at least one paper/article/blog relevant to your topic beyond the paper you are re-implementing/novel idea you are researching.
  - https://news.climate.columbia.edu/2023/01/12/what-uncertainties-remain-in-climate-science/
  - An article written by the Columbia Climate School titled "What Uncertainties Remain in Climate Science" explores how different external factors can lead to different outcomes in climate projection including sea ice prediction, making it harder to predict. These factors include how cloud formations and patterns can lead to warming or cooling, and natural variability, both of which can make it harder to predict longer-term climate change because of its short term effects on weather patterns. Thus climate models are inherently imperfect because there are so many external factors that we can't necessarily account for or accurately predict.
  - The 2023 paper "Forecasting Arctic Sea Ice Concentration using Long Short-term Memory Networks" similarly uses LSTMs to predict sea ice concentration over a 1-9 month period. One model forecasts sea ice concentration at specific individual grid points, and another model predicts sea ice concentration for all grid points at the same time. This paper uses 43 years of data, and its external predictors include accounting for atmospheric and oceanic variables.
  - In this section, also include URLs to any public implementations you find of the paper you're trying to implement. Please keep this as a "living list"–if you stumble across a new implementation later down the line, add it to this list.
  - https://dl.acm.org/doi/10.1145/3665053.3665054
  - https://dl.acm.org/doi/10.1145/3589883.3589901

    - https://dl.acm.org/doi/10.1145/3681765.3698457
- **Data**: What data are you using (if any)? If you're using a standard dataset (e.g. MNIST), you can just mention that briefly. Otherwise, say something more about where your data come from (especially if there's anything interesting about how you will gather it).
        - 1D LSTMs: https://nsidc.org/data/g02135/versions/3 obtained from the National Snow and Ice Data Center (NSIDC) and consists of monthly numerical data from 1989 to 2022. NSIDC monitors and provides data on various aspects of Earth's cryosphere, including sea ice, glaciers, and snow cover.
    - How big is it? Will you need to do significant preprocessing?
        - 1D LSTM: preprocessing techniques include normalization, dimensionality adjustments, and dataset merging. In addition to the SIE historical monthly values (target variable), we plan to collect five climate variables as control variables: sea surface temperature (SST), total precipitation (tp), surface pressure (SP), surface latent heat flux (SLHF), and surface sensible heat flux (SSHF). The ERA5 dataset contains the longitude, latitude, EXPVER, and time dimensions. EXPVER is a dimension representing the version or ensemble member of the experiment in climate or weather modeling. Our dataset includes 2 EXPVER values, indicating that there were two ensemble members or experiment versions in the subset of the NetCDF file. In addition to EXPVER, each climate variable in the NetCDF file will have its own 2D array of values representing the data for each combination of longitude and latitude. Our first step will be to split the dataset in two based on its EXPVER value. We will then conduct spatial aggregation to lower data dimensions based on latitude and longitude. This will effectively transform our dataset, enabling conversion to CSV format for merging with our numerical SIE.
- **Methodology**: What is the architecture of your model?
    - 1D LSTM model:
        - Activation function: sigmoid and tanh
        - Loss function: Mean Absolute Error
        - Optimizer: Adam
        - Consisted of multiple (78) LSTM cells
        - Batch size: 64
    - How are you training the model?
        - Using training dataset and tuning with validation dataset
    - If you are implementing an existing paper, detail what you think will be the hardest part about implementing the model here.
        - We think the most difficult part for the 1D model will be implementing the LSTM cell, as it has a few moving parts that we have not yet done in class.
- **Metrics**: What constitutes "success?"
    - What experiments do you plan to run?

    - For most of our assignments, we have looked at the accuracy of the model. Does the notion of “accuracy” apply for your project, or is some other metric more appropriate?
    - If you are implementing an existing project, detail what the authors of that paper were hoping to find and how they quantified the results of their model.
        - The paper we are re-implementing measures accuracy by comparing the predicted sea ice extent values of the 1D LSTM model to the actual sea ice extent values in the dataset. They use an R squared score, which is the coefficient of determination. The coefficient of determination will demonstrate how well the network is fitting the data, while the comparison of predictions to data will demonstrate accuracy.
        - Beyond the R squared score, we will also compare predictions to real data (something the paper did not implement)
    - What are your base, target, and stretch goals?
- **Ethics**: Choose 2 of the following bullet points to discuss; not all questions will be relevant to all projects so try to pick questions where there’s interesting engagement with your project. (Remember that there’s not necessarily an ethical/unethical binary; rather, we want to encourage you to think critically about your problem setup.)
    - What broader societal issues are relevant to your chosen problem space?
        - Climate justice is relevant to our chosen problem space because sea ice melting is not only caused by rising temperatures but also causes rising sea levels, further global warming, and shifting ocean circulation patterns. This is connected to human health, environmental health, and other social issues surrounding climate justice.
    - Why is Deep Learning a good approach to this problem?
        - Deep learning is a good approach to this problem as it is very data intensive. The sheer quantity of data makes it difficult to analyze manually to make predictions. Deep learning is an effective solution here as it can help make predictions on empirical data to inform decision-making.

- **Division of labor**: Briefly outline who will be responsible for which part(s) of the project.
    - We plan to meet together and work on the coding part of the project! :D
    - We will split up the write up/report/and presentation evenly but still collaborate throughout.