Reproducibility instructions:

Kaggle username: bluetriad, team name: blue
Score: 76.29 (Private), 84.33(Public)
Name: Ayush Mishra (ayushm3002@gmail.com)

In my submission I have included the .ipynb notebook I used for my submission, the notebook itself has links to the publicly available climate data that I have used. It also has my rationale behind my feature engineering methods and feature selection/elimination process.

My notebook can be run on any kaggle-like environment, I have attached kaggle's environment's docker file as well in my submission.

Lastly, to run my notebook with the external data, you can either download the different datasets from the public website link I have provided in my notebook or you can use the extra_climate_data.csv where all of it is combined and sorted according to the date. I have also hosted this file on a public kaggle dataset which is used in my kaggle submission.

I hereby confirm that I have used no future data to create my dataset or any external price data such as crude oil prices/ethanol prices as the spirit of this competition is to create features from climate data and my external data just adds to the available climate data.

The only time a future column is mentioned in my notebook is when they are used to calculate the CFCS Score for evaluation and feature selection/elimination.

Here is the basic schema of my feature generation pipeline.