Temperature Rising!

A UVA Data Science Case Study by Michael Vaden, 2022


Canada is warming, and very quickly. Since 1948, the average temperature in Canada has increased by over 1.7 degrees Celsius. Rain patterns are shifting and the sea level is rising. The Canadian government is concerned about the long-term effects of global warming on both the environment and the economy, and they have hired you (an up-and-coming data scientist) to analyze the effects of global warming.

Specifically, you need to create a collection of models to predict BOTH temperature and precipitation over time. Using the time series Canada dataset, you need to predict temperature and precipitation in Toronto. You should subset the data set to only include dates and Toronto data. You must use the dates in the range 2010-2019 as your test set, with all dates before as your training data. There are a plethora of predictive models available for use, but for this assignment you are required to implement at least one traditional machine learning approach as well as research and implement an approach used specifically for time series data (HINT: investigate models that consider seasonality). You may use the same traditional and time-series specific models for both temperature and precipitation in Toronto, but you must predict these values independently from one another.

When considering results of your models, you must use the same metric(s) of error across all of your approaches for comparison. You may choose whichever error measure you think is most appropriate, but you must explain why you chose that metric.

Deliverable: Give recommendation of two models, one of which is time-series specific, to predict temperature and precipitation respectively in Toronto. Make clear which error metric you used to evaluate your models, and include your code in your submission. Also include any relevant research materials that contributed to your final recommendations.

HINT: it is recommended that that you attempt and compare many different models before selecting two for your deliverable
BONUS: predict temperatures and precipitation for 2020-2029, and analyze the results