

1. How do you frame your main question as a machine learning problem? Is it a supervised or unsupervised problem? If it is supervised, is it a regression or a classification?

I am framing my main question as a question about machine learning as a look at how all the individual factors that I am tracking (Snow Water Equivalent, Atmospheric CO₂, visitor numbers and tourism income) interact and are correlated to each other.

This is a supervised problem as I am running linear models on many data frames including frames in which I am comparing two variables against one another as well as tracking the residuals of the variables and comparing those in a linear model.

This is a regression, as I am tracking the regression of these models and forecasting the future of what will happen with these models.

2. What are the main features (also called independent variables or predictors) that you'll use?

The main independent variable I am looking at is atmospheric CO₂ concentration. This is because the main idea of my capstone is that the increase in atmospheric CO₂ causes lower snowfall recorded in Snow Water Equivalent and therefore it also causes lower tourism and income due to the correlation of lower snowfall and visitors, and therefore income.

3. Which machine learning technique will you use?

I will be using linear regression as the machine learning technique in my capstone project.

4. How will you evaluate the success of your machine learning technique? What metric will you use?

I will test the validity by checking residuals and checking the P values of my linear regressions using the summary function of my linear models. I will check residuals using the checkresiduals function.