Predicting Cherry Blossom Bloom Dates Using Elastic Net

Cherry blossom trees are an iconic symbol of spring's arrival and hold cultural and symbolic significance in many parts of the world. They are widely celebrated in Japan and the United States. People travel from all over the country to observe the fleeting cherry blossoms and attend parades and festivals in their honor. The fragile pink and white flowers mark the end of winter and the beginning of spring. They represent hope, birth, and renewal. In addition to their cultural and symbolic significance, cherry blossoms are a key indicator of climate change in tree phenology, as temperature is the primary driver of phenological development. However, predicting bloom dates is a notoriously challenging task. This is due to various factors, including weather patterns, temperature fluctuations, and other environmental factors.

Cherry blossoms require a delicate balance of warm and cold temperatures. During the winter, they need to acquire a certain amount of chilling accumulation to transition them out of endo-dormancy so they can begin to grow. Then, they must accumulate a certain amount of heat to induce blossoms. The amount of chilling accumulation and heat accumulation required is largely unknown. Thus, I aimed to predict cherry blossom bloom dates using an elastic net model trained on multiple temperature-based variables.

Growing degree days (GDD) were used to measure heat accumulation. GDD was calculated using the Average Method and the Baskerville-Emin (BE) Method. The Average Method is the most common. Yet, the literature suggests that the BE Method is superior when the minimum temperature is below the base temperature, which is most likely to occur in the spring. Thus, both methods were included in the model. To measure chill accumulation, chill days (CD) were computed by summing the number of days where the maximum temperature fell below 7.2 ℃. Quetlet’s law of flowering plants was also used, where temperature is substituted for time in Newton’s second law of motion. This theory suggests that plants bloom when the sum of the mean daily temperature squares reaches a certain threshold. Accordingly, I included the sum of squares and cubes for each predictor (Average GDD, BE GDD, and CD) in the model. Due to the clear multicollinearity issue between the predictor variables, elastic net regression was used. This modern method allows users to include correlated predictors because it includes both variable selection and regularization, which are respectively found in lasso and ridge regression.

80% of the data was used to train the model, and 20% was used to test the model. The model was cross validated with five folds. The R-squared was 0.421, indicating that the model explains 42.1% of the variation in cherry blossom bloom dates. The RMSE was 7.15, indicating that the model, on average, predicts the bloom date within a week (7 days). The model was compared to a simpler elastic net regression model that only included the Average GDD, BE GDD, and CD variables. The simpler model explained only 25.99% of the variance in bloom dates, indicating that the additional variables used in the original model could explain an extra 16.11% of the variance.

In conclusion, predicting cherry blossom bloom dates is challenging due to various environmental factors affecting their development. However, the elastic net model used in this study was able to account for multicollinearity between predictor variables and explain a significant amount of the variance in bloom dates. By better understanding the factors that affect cherry blossom tree phenology, we can gain valuable insights into the effects of climate change on these important cultural and symbolic symbols. However, further research is needed to better understand the complex relationship between environmental factors and cherry blossom tree phenology, particularly in the context of climate change.