1 Introduction

Many cities are known for their cherry blossom viewing sites, and every year, thousands of tourists visit those places to watch cherry blossoms. However, many tourists on social media complained that either they went to the viewing sites, but the cherry had not bloomed, or the cherry had already started to wither. Two of our team members had this experience themselves. Thus, knowing the cherry blossom dates is beneficial for the tourists to arrange their trips, helping them save money and time.

In this project, we explore the 6 predictors of daily average temperature, humidity, pressure, wind speed, daily precipitation, and effective accumulated temperature, utilizing various models to predict cherry blossom dates at Washington, D.C. (USA), Kyoto (Japan), Liestal-Weideli (Switzerland) and Vm dates are beneficial for the tourists to arrange their trips, helping them save money and time Vancouver, BC (Canada).

2 Data

Instead of the cherry blossom date, we use blossom day (from Jan 1 of each year) as our dependent variable. Using JavaScript, we also scraped weather data such as daily average temperature, humidity, pressure, wind speed, and daily precipitation from The Weather Channel. Inspired by the paper[1], we created a new variable called effective accumulated temperature, which measures the sum of maximum temperatures of each day from a predefined starting date to peak blossom date[2]. To determine the starting date, we calculated the mean absolute value error of multiple starting dates and selected the starting date with the lowest error. When the effective accumulated temperature of cherry blossoms reaches the accumulated temperature threshold, the corresponding data is the peak blossom period of the cherry trees. For each of the 4 cherry blossom locations, the temperature thresholds and optimal starting dates are different due to local climate and cherry tree species. Especially, due to the lack of cherry blossom days in Vancouver, we cannot calculate the corresponding start date and temperature threshold. The results are shown below:

Table 1: Start Date and Temperature Threshold

Region	Start Date (doy)	Temperature Threshold ($^\circ C$)
Washington	30	712.431
Kyoto	58	563.227
Liestal	25	673.889

The variables in our models are listed as follows:

Table 2: Variable Description

Variable	Definition
Year	Year from 1945 to 2023
Date	Day of the month
Temperature	We have both yearly average temperature and daily average temperature in our dataset.
Max Temperature	Daily maximum temperature
Accumulated Temp	Sum of maximum temperatures of each day from predefined starting date to peak blossom date
Precipitation	Any product of the condensation of atmospheric water vapor that falls under gravitational pull from clouds. We have both yearly precipitation and daily precipitation in our dataset.
Humidity	The concentration of water vapor present in the air. We have both yearly humidity and daily humidity in our dataset.
Pressure	Air pressure, also known as atmospheric pressure, is the pressure within the atmosphere of Earth. We have both yearly air pressure and daily air pressure in our dataset.
Wind Speed	A fundamental atmospheric quantity caused by air moving from high to low pressure, usually due to changes in temperature. We have both yearly wind speed and daily wind speed in our dataset
Blossom Day	The peak blossom day counting from Jan 1st each year

3 Methodology

Washington DC and Liestal
To predict Washington's data, we experimented with several models, including random forest, support vector machine (SVM)[3], time series autoregression, long short-term memory (LSTM), and neural network. We also utilized linear regression model and considered many instrumental variables.
Kyoto
The Kyoto dataset contains historical cherry blossom data dating back to the Year 812. However, due to limited temperature data we collected, we considered data from the Year 1997. We agreed that analyzing long-term climate trends requires considering time-dependent patterns in the data. And time series techniques, such as ARIMA modeling, are commonly used to model such patterns and predict future observations. Therefore, we believed ARIMA modeling is the most suitable method for analyzing the Kyoto dataset.
Vancouver
We barely have any data related to Vancouver. So we used the "USA NPN" dataset as a training set. For the Vancouver site, we included the differences between the latitude, longitude, and altitude of Vancouver and US cherry viewing sites. Also, we utilized AGDD as one of our predictors. Among all the prediction models we tried, Random Forest is the best model.

4 Results

The Accuracy of each model is shown below:

Table 3: Model Accuracy

City Name or Algorithm	Washington	Kyoto	Liestal	Vancouver
SVM	0.4	0.4	0.16	N/A
Linear Regression	0.40	0.33	0.33	N/A
Logistic Regression	0.47	0.47	0.27	N/A
Random Forest	0.73	0.67	0.67	0.53
Cumulative Temperature	N/A	N/A	N/A	N/A
ARIMA	0.4	0.73	0.4	N/A

5 Conclusion

Every machine learning model worked differently on different datasets. Random Forest and ARIMA stood out from all the models we tried, and we used them to predict the blossom date in the future 10 years. However, our maximum accuracy was 0.73, which is not high in the machine learning field. For future experiments, we aim to tune the model better or adjust the model based on this scenario.

References

[1] Chen Chao, Tian Shan, Li Mei-Yan, and Xiao Xiao. The prediction model of initial flowering based on accumulated temperature of different species peach in longquanyi. *Proceedings of the ACM Turing Celebration Conference - China*, 2019.

[2] Pleases Miss. 2020 japan cherry blossom season travel organization: How to scientifically predict the flowering time, cherry blossom full bloom prediction, and must-check cherry blossom viewing frontline sites. *Cool3c*, Jan 2020.

[3] Hongyi Zheng, Yanyu Chen, and Zihan Zhang. Predicting blossom date of cherry tree with support vector machine and recurrent neural network. *arXiv:2210.04406v1*, Oct 2022.