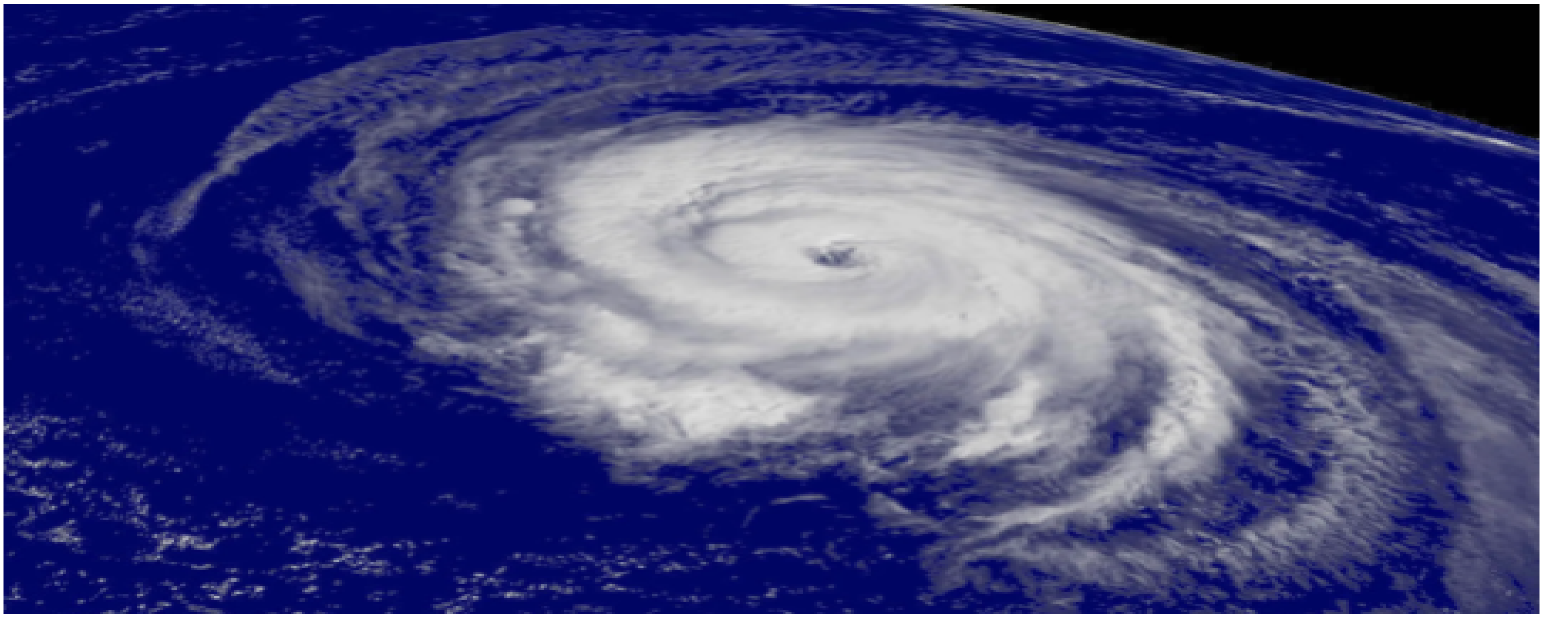


New Jersey Institute of Technology

CS675: Group Project, Milestone 4

Alexander Guillen

Doug Rizio

William Duggan

Table of Contents

Introduction

The subject of climate change is everywhere. From scientific studies and news headlines to popular media and even casual conversations, the human collective consciousness is dominated by concerns of the world's climatic conditions – and for good reason, too. According to the United Nations Security Council, climate change is the "biggest threat modern humans have ever faced."

If we don't do something to stop our carbon output now, the planet faces catastrophic consequences – more frequent and intense droughts and storms, rising sea levels, melting glaciers, warming oceans, the destruction of habitats, the degradation of crops, human migration on an unprecedented scale, and social unrest so great that it threatens to destabilize civilization itself.

Ultimately, most of these problems stem from one major cause: global warming. In order to further understand how global warming has developed over the course of recent human history, some of the best sources of information are datasets on global temperature changes over time. The FAOSTAT Temperature Change dataset provides exactly this kind of resource, measuring temperature in each country of the world from 1961 to 2021.

Abstract

Our group was really excited to discuss and explore this dataset as we have members in the environmental industry and are passionate about climate change data. We decided to divide into 3 sub-sections and each explore the different regions of the globe.

After evaluating our models, there were some similar conclusions, of which, to be discussed. We also wanted to see how far we could explore by predicting what climate change could look like in the near future.

Source URL

The following link will take you right to the website where you can pull "almost" live datasets to analyze. Our group downloaded our own tailored csv files to work with.

FAOSTAT Climate Data

Understand Dataset

The Global Surface Temperature Change data distributed by the National Aeronautics and Space Administration Goddard Institute for Space Studies (NASA-GISS) is publicly available. The FAOSTAT Temperature Change domain includes some extra information in this data set like Area names. Because in this project our key value is country names, we chose to use FAO's data set.

Data in the Temperature Change domain can be reachable from the Food and Agriculture Organization of the United Nations web data portal. According to the license of FAO Statistical Database Terms of Use, the data set can be used for research, statistical, and scientific purposes. It can be accessed, downloaded, created copies and re-disseminated datasets subject to these Dataset Terms.

The FAOSTAT Temperature Change domain disseminates statistics of mean surface temperature change by country, with annual updates. The current dissemination covers the period 1961–2021. Statistics are available for monthly, seasonal and annual mean temperature anomalies, i.e., temperature change with respect to a baseline climatology, corresponding to the period 1951–1980. The standard deviation of the temperature change of the baseline methodology is also available.[2] It includes areas of all the countries and territories of the world. The data covers monthly, seasonal, yearly temperature changes as Celsius degrees °C. The frequency of dissemination and release calendar of the data is the yearly base. The format is a 3 separate comma-separated value (CSV) files, and has the tabular format and totals 67.9 megabytes.

Further examination on the attributes and features revealed years with some missing values. However, we will attempt to visualize in depth using most key values provided below;

- Area Code: The numerical code of area column, type of area code is an integer.
- Area: Countries and Territories (In 2019: 190 countries and 37 other territorial entities.), type of area is an object.
- Months Code: The numerical code of months column, type of months code is an integer.
- Months: Months, Seasons, Meteorological year, type of months is an object.
- Months: 'January', 'February', 'March', 'April', 'May', 'June', 'July','August', 'September', 'October', 'November', 'December'
- Seasons: "Dec\Jan\Feb", "Mar\Apr\May", "Jun\Jul\Aug","Sep\Oct\Nov'
- Year: 'Meteorological year'
- Element Code: The numerical code of element column, type of element code is an integer.
- Element: 'Temperature change', 'Standard Deviation', type of element is an object.
- Unit: Celsius degrees °C, type of unit is an object.

Section 1: "The Americas"

The Americas region was interesting to dissect. It was very profound to see the difference in temperatures across seasons, decades and months. While remembering what it has been like the past 20 years living in the states, it was interesting to see that our intuition aligned with the data trends,

There were some recent years where it felt that December, January, and February were not particularly "cold" and even had some holiday seasons without much snowfall. It just so happens that these months resulted in the most significant changes from other seasons.

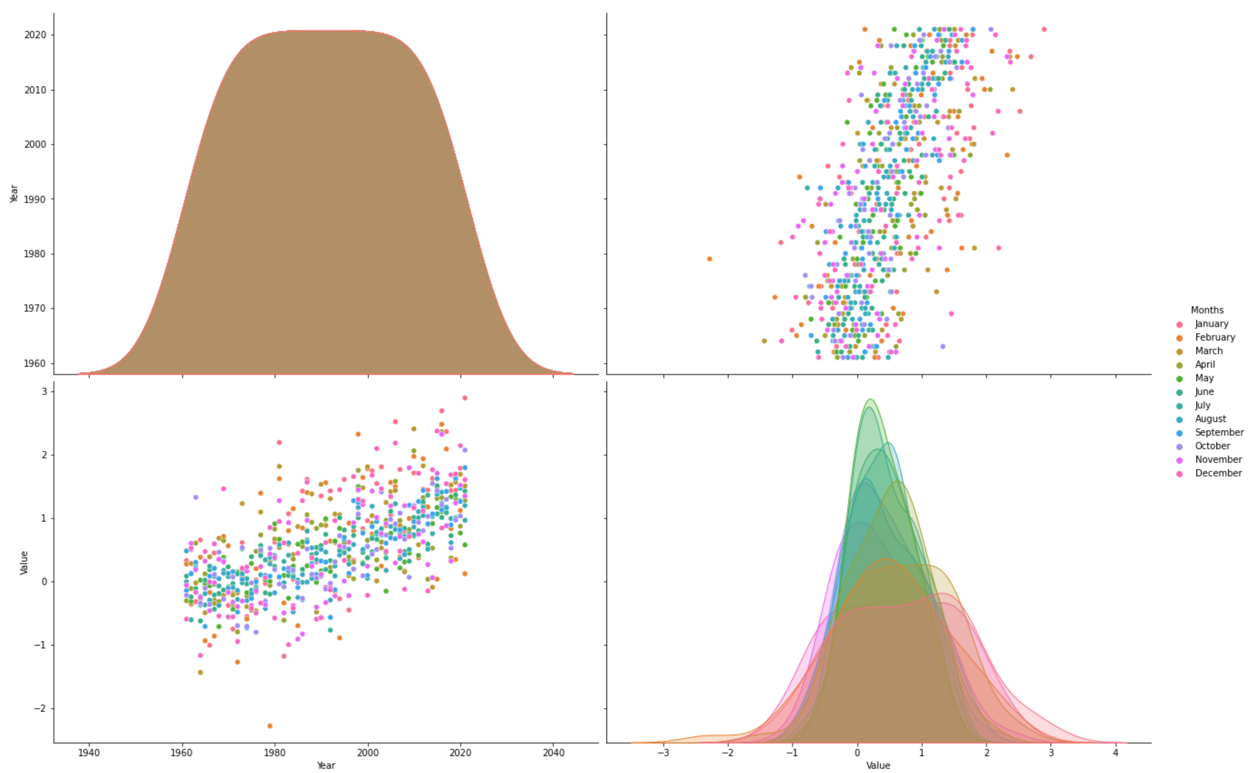


Figure 1.1 Pairplot illustration depicts the temperature change over the months across the decades.

There were trends found in specific months that were observed. Breaking down subsets of the original dataset became more frequent in respect to latter years that also highlighted the increase in Temperatures and Standard Deviations. Slicing the metrics by colder months, and seasonally, allowed the data to illustrate these changes as well.

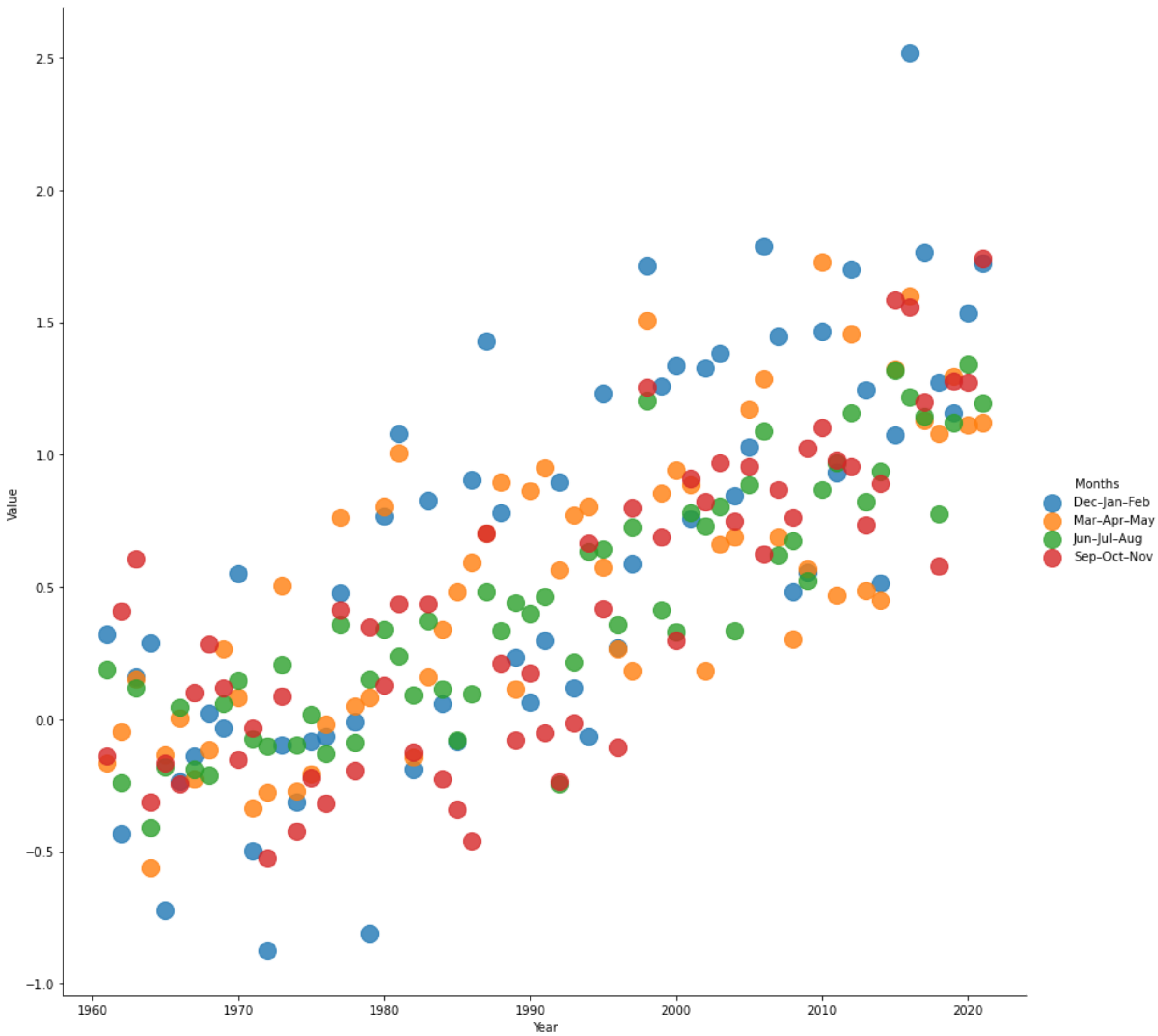


Figure 1.2 Scatterplot of seasonal data

KNN was the best model to score this dataset as our questions were more towards classification to compare which subsets could be identified according to the change in temperatures over the years. There were a few outliers detected, with recent spikes found for the “Dec-Jan-Feb” grouping. This was more noticeable and visually represented upon initial review.

Algorithm	Model	Adaptivity Level	Accuracy	Precision	Recall	F1_Score
Logistic Regression	MultiNomial	Low	68%	0.75	0.62	0.68
		Moderate		0.64	0.82	0.72
		High		0.71	0.31	0.43
Naive Bayes	MultiNomialNB	Low	68%	0.77	0.60	0.68
		Moderate		0.63	0.82	0.71
		High		0.65	0.38	0.48
	ComplementNB	Low	64%	0.70	0.66	0.68
		Moderate		0.61	0.71	0.66
		High		0.55	0.28	0.37
	CategoricalNB	Low	70%	0.76	0.63	0.69
		Moderate		0.66	0.79	0.72
		High		0.69	0.56	0.62
Support Vector Machine	SVC	Low	89%	0.93	0.89	0.91
		Moderate		0.86	0.93	0.89
		High		0.84	0.69	0.76
	LinearSVC	Low	67%	0.75	0.59	0.66
		Moderate		0.62	0.82	0.70
		High		0.75	0.31	0.44
Neighbors	KNeighborsClassifier	Low	90%	0.93	0.91	0.92
		Moderate		0.87	0.93	0.90
		High		0.87	0.69	0.77

While the visualizations clearly favor the rise in temperature over the decades, we wanted to explore possibilities of predicting the future and what climate may look like. In the next section, the dataset, and additional machine learning tools, allow us to explore this further.

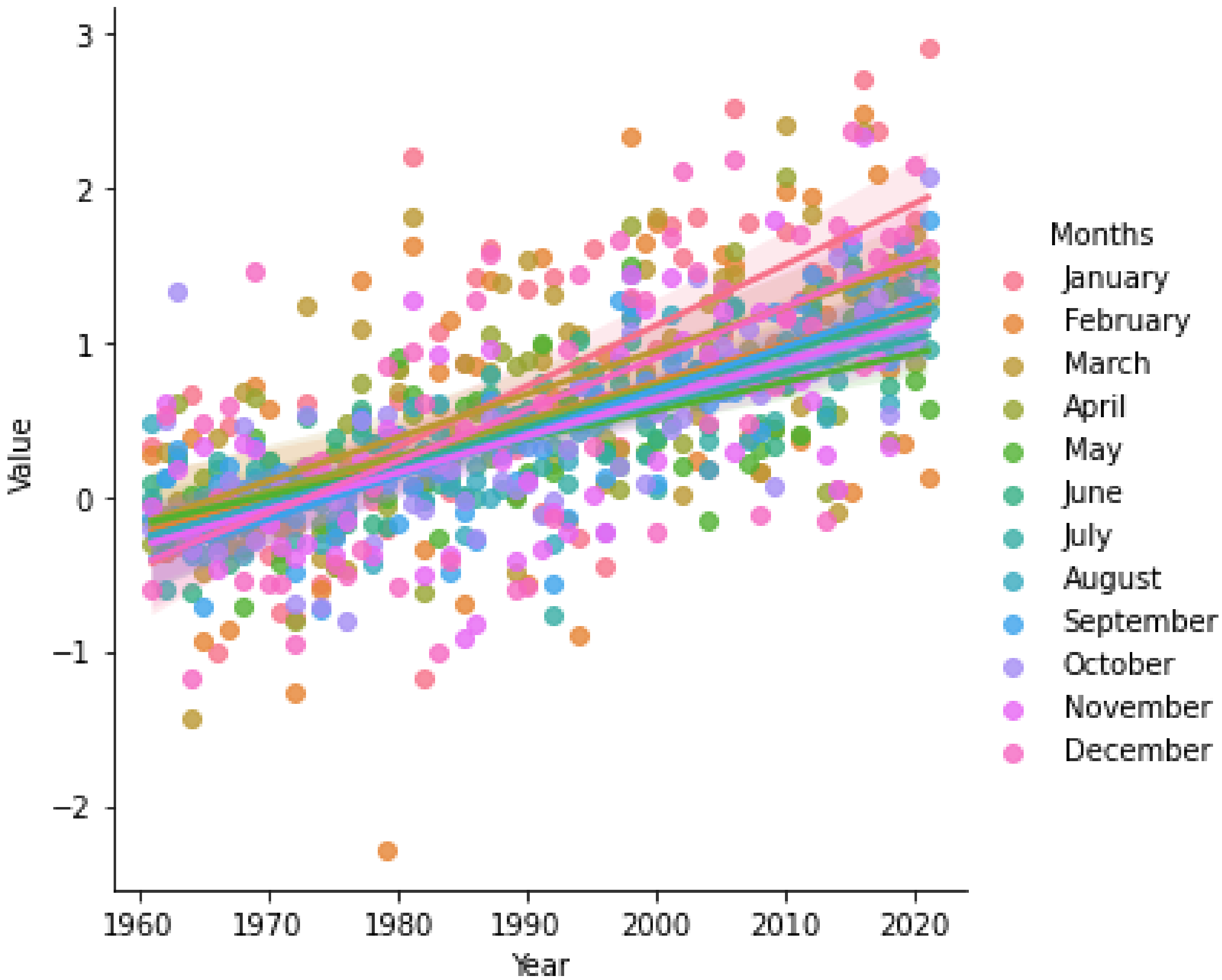


Figure 1.3 Months and years by value.

Section 2: "Europe"

To visualize the temperature change in Europe, the mean temperature change (1961 – 2021) was obtained from https://www.fao.org/faostat/en/#data/ET/visualize[1]. Figure 2.1 corresponds to the mean temperature change by decade. It appears that the temperature change tends to fluctuate throughout the year as expected. Further, the data suggests that the temperature in Europe has risen steadily over the past few decades.

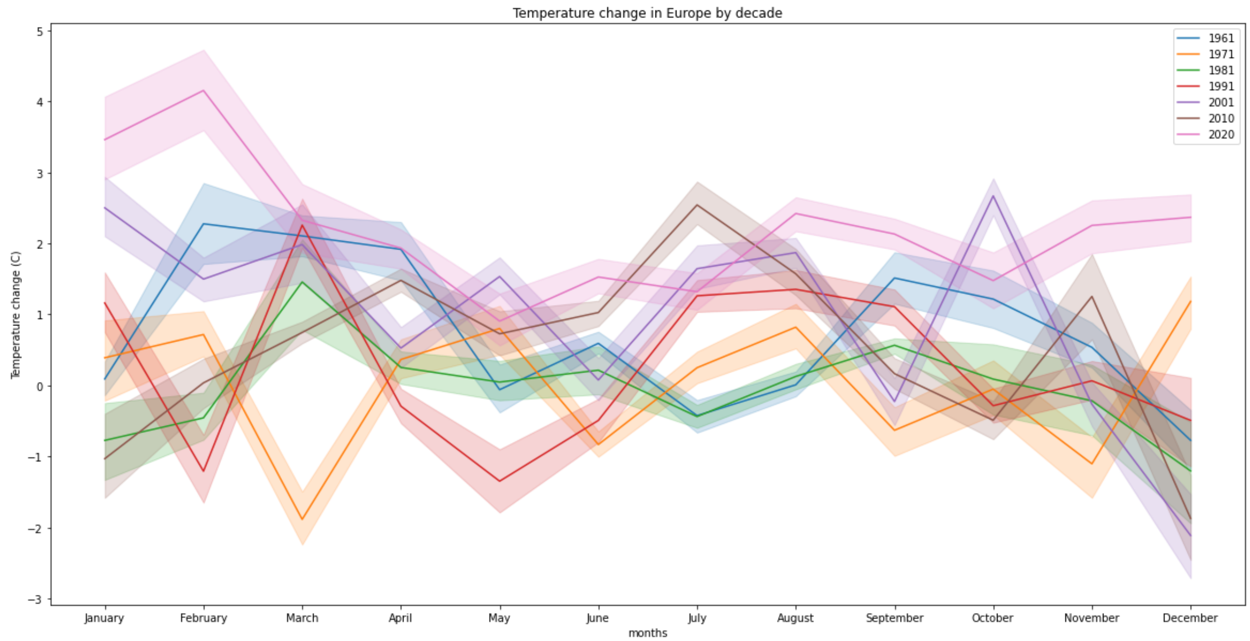


Figure 2.1 The mean temperature change by decade (1961 – 2020)

To predict the temperature change in Europe, two different approaches were used: Prophet and Random Forest Regressor.

Prophet, or Facebook Prophet is an open-source library for univariate time series forecasting developed by Facebook [2]. Prophet implements what they refer to as an additive time series forecasting model, and the implementation supports trends, seasonally and holidays [2]. This model was designed to be easy and completely automatic, and it is typically used for forecasting sales, capacity, etc. [2].

There are many sources and examples available online for the implementation of this algorithm. A modified version of the code obtained from 📈 Global Temperature Time Series Forecasting | Kaggle[3] was used for the prediction of the temperature change in Europe. Implementing this algorithm requires three steps: pre-processing, in-sample forecast, and out-of-sample forecast.

In the pre-processing step, the unwanted information was removed from the dataset, and the mean temperature change was obtained, as shown in **figure 2.2-A**, validating the results obtained in **figure 2.1**. Then, *in-sample forecasting* (**figure 2.2-B**) was performed to train, test, and fit the model. Although the model created with Prophet showed a similar trend to the results obtained in figure 1 and figure 2.2-A**---**the average temperature increases over time **---** it fitted the actual data very poorly. Finally, *out-of-sample forecast* (**figure 2.2-C**) was made to make predictions beyond the training data. Like figure 2.2-B, the model predicted using the Prophet algorithm does not seem to show reliable results.

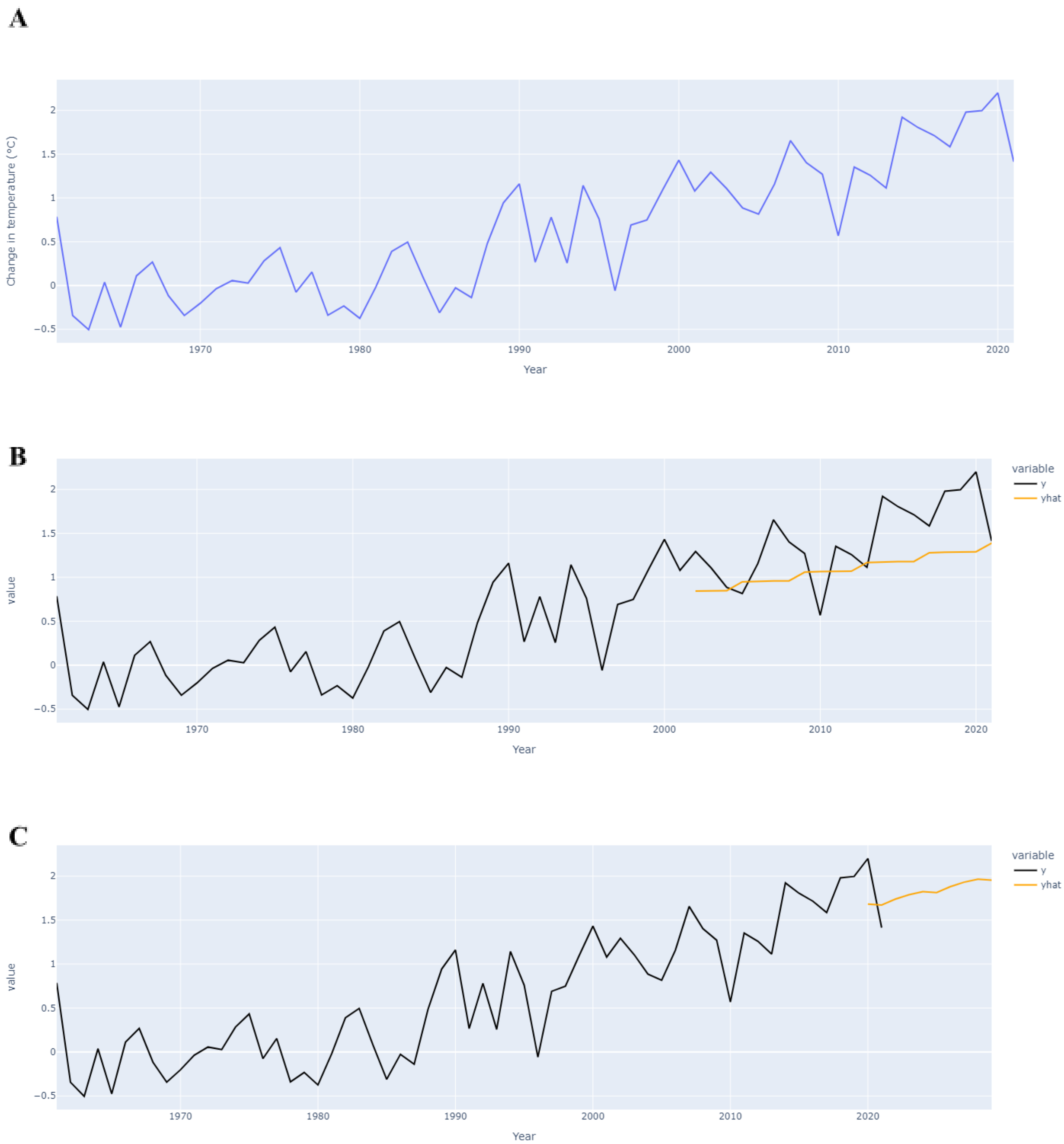


Figure 2.2 Average European Temperature change (1961-2021). **A** corresponds to the pre-processing step. **B** shows the results obtained from the in-sample forecasting step, and **C** is the out-of-sample forecasting results.

It is evident from the figures above that making predictions with Prophet does not necessarily give accurate results. Hence, a major criticism against Prophet is that its underlying assumptions are simplistic and weak [4]. It does not look for casual relationships between the past and the future. It simply finds the best curve to fit the data using a linear logistic curve component for the external regressor [4].

The Random Forest Regressor algorithm was then used to try to make better predictions. This model is a supervised learning algorithm that uses ensemble learning method for regression [5]. Ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model [5]. Four steps were used to implement this algorithm: (1) identify the dependent and independent variables, (2) split the dataset into the training set and test set, (3) training the random forest regression on the whole dataset, and (4) predicting the test results. Figure 2.3 shows the graphic of the actual values versus the predicted values. Using the model, initially the predicted values do not quite match the actual values, but over time the values seem to fit the curve of the actual values relatively well.

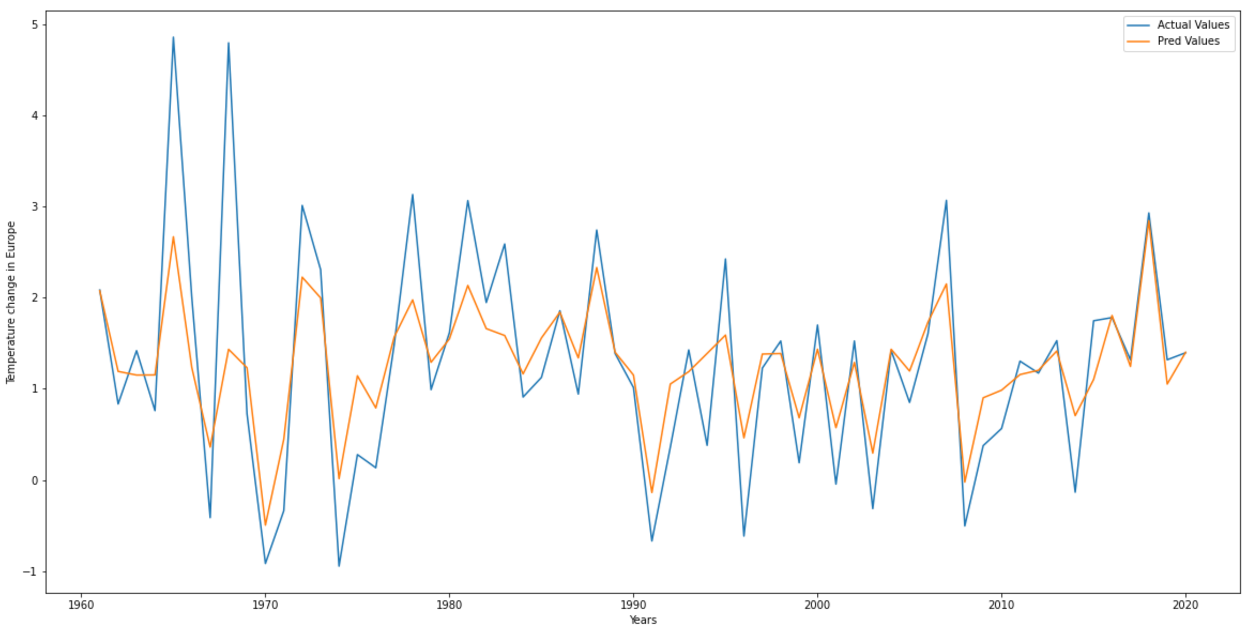


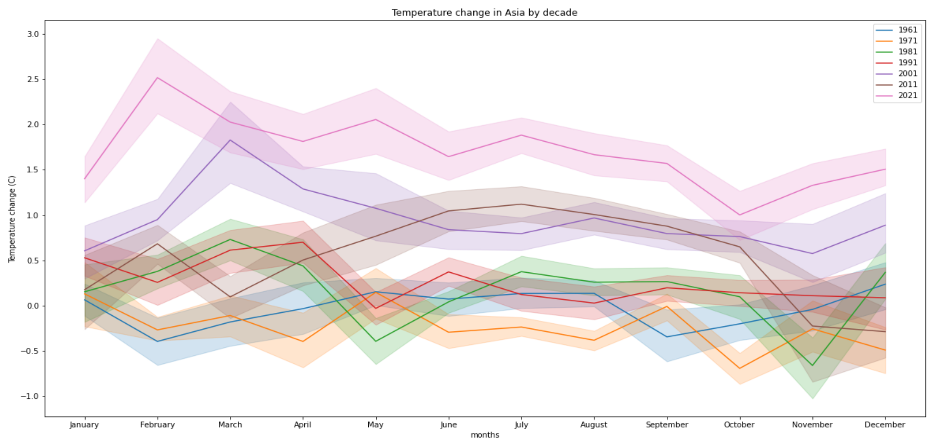
Figure 2.3 Temperature Change in Europe using Random Forest Regressor

Section 3: “China, Oceania and the West”

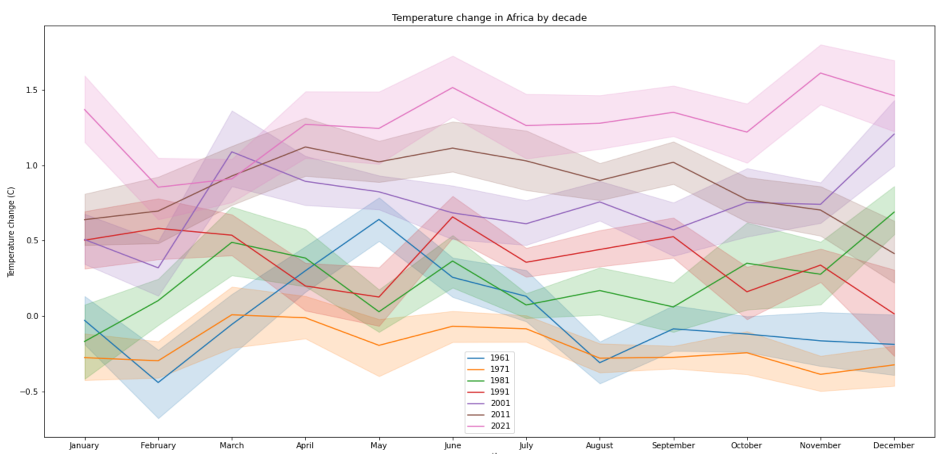
Visualizing data from Asia, Africa, and Oceania using the same methods as above, it is apparent that all four regions have very different temperature change profiles compared to each other.

Whereas in Europe the overall range of temperature change that occurs from 1961 to 2021 is between roughly -3.0 and +5.0 degrees celsius (an 8.0 degree range), the overall range of temperature change throughout the other regions is from only -1.0 to -3.0 for Asia (4 degrees) as shown in **figure 3-A**, -0.5 to 1.5 for Africa (2 degrees) as shown in **figure 3-B**, and -0.5 to 1.5 for Oceania (2 degrees again) as shown in **figure 3-C**. While these differences may be partially explained by the already high seasonal variability of countries in Northern latitudes (European countries especially), they may also suggest that these regions are experiencing temperature change at an accelerated rate compared to their counterparts across the globe. However, one limitation of this analysis is that these regions of the world are extremely diverse, and feature a wide variety of subregions with latitudes and climactic conditions that could act as outliers which push the minimum and maximum temperatures towards greater extremes.

A



B



C

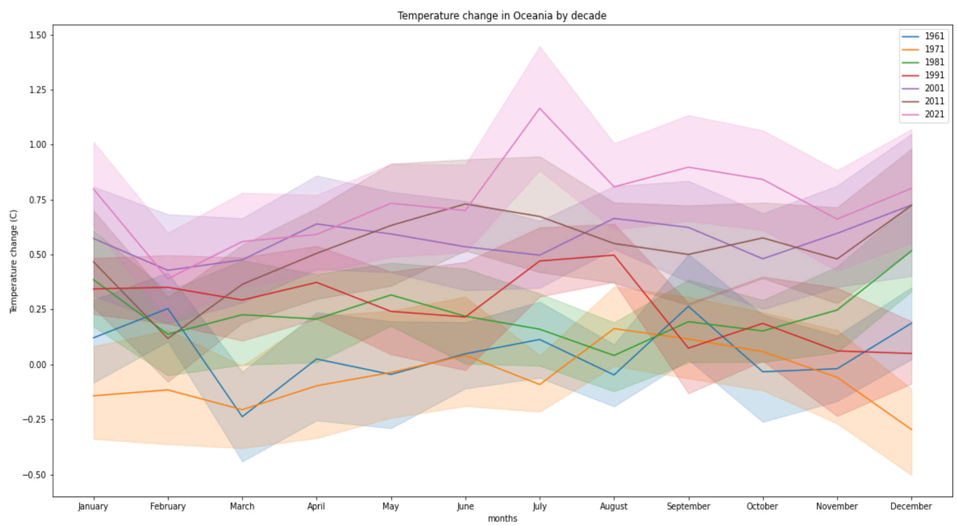
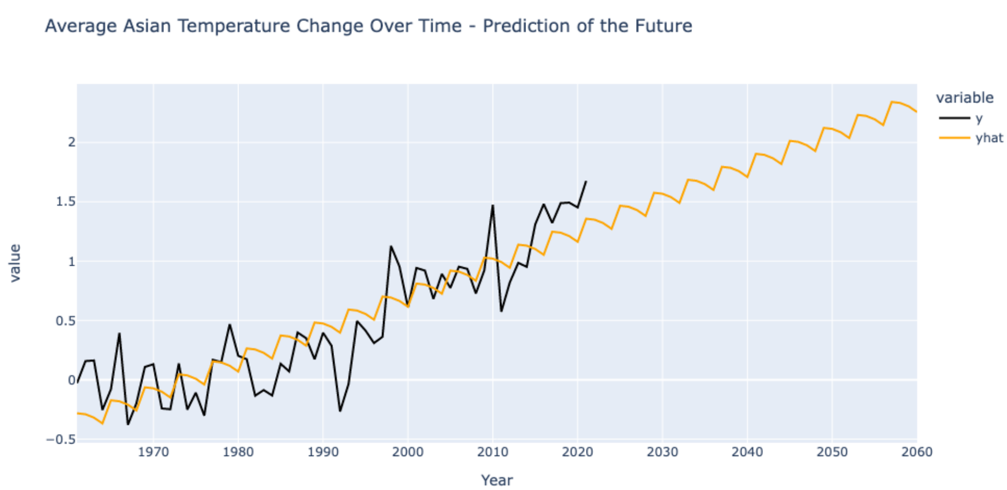


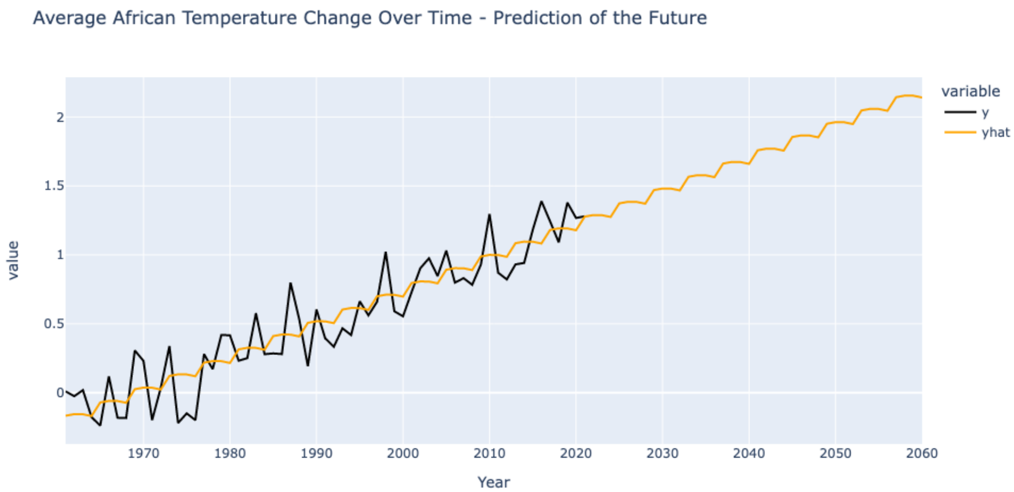
Figure 3. Average Asian, African, and Oceanian Temperature change (1961-2021). **A** shows the data for Asia, **B** shows the data for Africa, and **C** shows the data for Oceania.

And, while it was already established that the Prophet algorithm was not necessarily the most accurate predictor for this dataset (Europe in particular), it was still worthwhile to run it on the other regions of the world to compare the various predictions to each other. For both Asia (**figure 4-A)** and Africa **(figure 4-B)**, the total temperature change from 1961 to 2061 is predicted to reach almost 2.5 degrees, whereas for Oceania the total temperature change is predicted to reach just under 1.5 degrees **(figure 4-C)**. However accurate the model, this further illustrates that while temperature change is happening everywhere, every region is experiencing it differently, and the ultimate fate of each region as it relates to climate is unique.

A



B



C

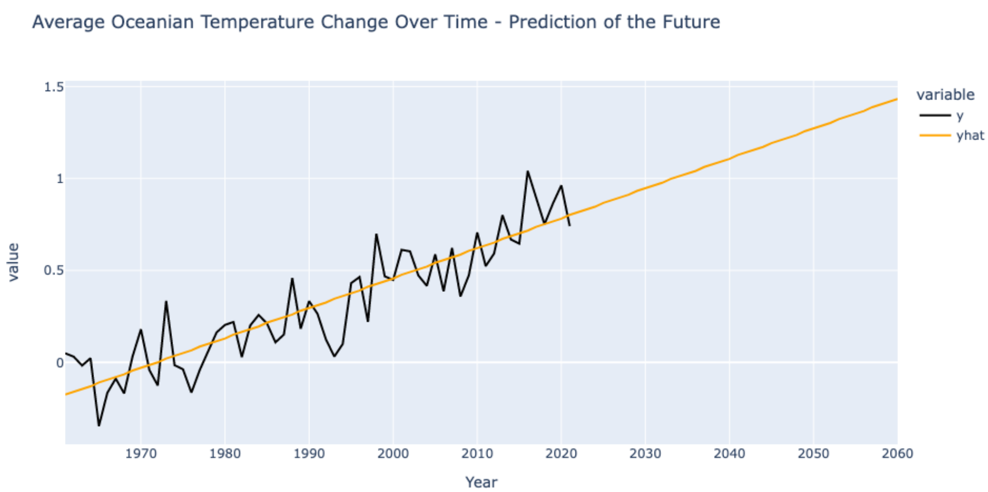


Figure 4. Predicted temperature change of the future for Asia, Africa, and Oceania (1961-2021). **A** shows the data for Asia, **B** shows the data for Africa, and **C** shows the data for Oceania.

Conclusion

"On July 19, the U.K. obliterated its all-time temperature record with a high of 104.5 degrees Fahrenheit. That bested the previous record of 101.7°F, or 38.7°C, which was set in July 2019. That staggering heat, in a country unaccustomed to it, came amid a 2°F (1.2°C) rise in global average temperatures since mankind began pumping greenhouse gas emissions into the atmosphere at the start of the Industrial Revolution.

Using computer climate modeling and analyzing data from weather stations throughout the U.K., researchers with the World Weather Attribution initiative determined that "the likelihood of observing such an event in a 1.2°C cooler world is extremely low, and statistically impossible in two out of the three analyzed stations [7]."

We experimented with various factors for different classification methods to get different results from the kaggle repository which was used. We also created models using Tensorflow, Prophet, and many more ML libraries.

To better use this data set we can also perform more data exploration and analysis to find :

- With the given data, we can classify the data sets by location , explore where the problems are and identify an action plan. The more people are aware of the data, the more they can understand that it is not going away and that we need to make a change now.
- We can find the target audience to talk about sustainable adaptation to this type of learning, appoint industry correction plans to be held accountable and begin making a real difference before it becomes too late.
- If we do not act now, the lifestyle our kids will live will be in uncomfortable, and even hazardous conditions.

References:

1. Security Council. (2021, February 23). *Climate Change 'Biggest Threat Modern Humans Have Ever Faced', World-Renowned Naturalist Tells Security Council, Calls for Greater Global Cooperation.* United Nations. https://press.un.org/en/2021/sc14445.doc.htm

2. FAOSTAT. (n.d.). *Temperature Change*. Faostat. Retrieved July 31, 2022, from https://www.fao.org/faostat/en/#data/ET/visualize

3. Brownlee, J. (2020, May 8). *Time series forecasting with Prophet in python*. Machine Learning Mastery. Retrieved July 31, 2022, from https://machinelearningmastery.com/time-series-forecasting-with-prophet-in-python/#:~:text=Prophet%2C%20or%20%E2%80%9CFacebook%20Prophet%2C,trends%2C%20seasonality%2C%20and%20holidays.&text=%E2%80%94%20Package%20'prophet'%2C%202019.

4. Gcdatkin. (2021, January 26). 📈 *global temperature time series forecasting*. Kaggle. Retrieved July 31, 2022, from https://www.kaggle.com/code/gcdatkin/global-temperature-time-series-forecasting

5. Goled, S. (2021, October 19). *Why are people bashing Facebook Prophet*. Analytics India Magazine. Retrieved July 31, 2022, from https://analyticsindiamag.com/why-are-people-bashing-facebook-prophet/#:~:text=A%2

0major%20criticism%20against%20Prophet,assumptions%20are%20simplistic%20and%20weak.&text=In%202017%2C%20Facebook%20released%20Prophet,outpaces%20the%20analysts%20producing%20them.

6. Bakshi, C. (2022, April 14). *Random Forest regression*. Medium. Retrieved July 31, 2022, from https://levelup.gitconnected.com/random-forest-regression-209c0f354c84
7. World Weather Attribution; Feed" ref="https://www.worldweatherattribution.org/feed/"