Social and Political Factors of Climate Change

Darren Zheng darrenz@princeton.edu
Sara Schwartz sarats@princeton.edu
Brian Lin bylin@princeton.edu

Abstract

The human impact on climate change has been undeniable. As such, we cannot afford to ignore the social and political aspects of this issue. In this assignment, we primarily focus on exploring the relationship between the American people's opinions on climate change and how they relate with greenhouse gas emissions and political party affiliation. Specifically, we build regression models to examine how climate change opinion relates with political party affiliation and to predict climate change opinion from greenhouse gas emissions and temperature. Our analysis primarily focuses on stratifying and analyzing each state in America. To predict climate change opinion, we find that random forest performs the best and SVM performs the worst. To predict political party affiliation, we see that linear regression performs the best and random forest performs the worst.

1 Introduction

Our climate impacts almost every aspect of our lives, such as our agriculture and food security, water and energy supply, transport infrastructure, and health. In the past few decades, our high rates of greenhouse gas emissions and over-exploitation of natural resources have drastically increased the natural timeline of global warming, posing an imminent threat to the quality and maintainability of life on our planet. While there has been a massive push in both legislative and individual actions we can take to slow (and hopefully, stop) climate change, climatologists warn that we are not moving fast enough to avoid some of the terrible consequences of climate change on our quality of life.

For our project, we wished to explore the relationship between real climate and greenhouse gas emissions data and the sociopolitical opinions surrounding global warming. In order to standardize the metric of our investigation, we decided to focus on each state's contribution to global warming through greenhouse gas emissions, their opinions on climate change, and their political party affiliation. First, in section 1, we build regression models to predict average temperature based on annual greenhouse gas emissions for each state. Next, in section 2, we visualized the similarities between states and counties in terms of climate change opinion through graph construction. We then built regression models to predict the percent of each state's population that affiliated with the two major political parties based on the attitudes of the state's citizens regarding climate change opinions. Finally, to tie it all together in section 3, we explored the relationship between each state's contribution to climate change (their greenhouse gas emissions and temperature data) and their opinions on climate change. We believe that these questions add to the value of current work using machine learning for climate change by providing a new source of information for scientists, educators, and policymakers to more effectively address how we - from the individual level to the national and international level - can more effectively address the challenges that climate change is imposing on our current and future way of living.

2 Related Work

Currently, the most common application of machine learning to topics surrounding climate change appears to focus on the relationship between greenhouse gas emissions and global

temperature. One example is a study done using 1.6 billion temperature reports from The Berkeley Earth Surface Temperature Study to build a prediction model for global surface temperature up to 2050, based on the average Carbon Dioxide in the atmosphere per year [1]. So, we first decided to build prediction models to investigate the relationship between greenhouse gas emissions and global temperature, and see if we can produce similar results to these studies.

However, our goal for this project was not simply to visualize the effects of climate change; we were also interested in investigating the social and political nature of this global issue. We were inspired by an article from the Nature journal which examined geographic variation in opinions on climate change at state and local scales in the USA [2]. Using a multilevel regression and post-stratification model, the article accurately predicted climate change beliefs, risk perceptions and policy preferences using a concise set of demographic and geographic predictors. To our surprise, the article found that nationally, only 63 percent of Americans believe global warming is happening. We wished to take this analysis one step further by investigating the relationship between each state's average climate change opinions and the political party the majority voted for in the 2012 and 2016 elections. Additionally, we combined the first article's data on each state's greenhouse gas emissions with the second article's data on each state's climate change opinions to determine if there is a relationship between state opinions on climate change versus their contributions to climate change.

3 Section 1: Predicting Temperature Based on Emissions

3.1 Data Processing

We obtained data on greenhouse gas emissions from 1990 to 2018 for 49 states excluding Hawaii from the U.S Environmental Protection Agency. As we wished to conduct analysis state-by-state, for each state, we scaled all of the corresponding values (such as 'greenhouse gas emissions' and 'electric power') by the state's total population. By doing so, we will be able to fairly compare each state's contribution to greenhouse gas emissions and global warming and avoid making skewed comparisons between two states with vastly different populations. This data serves as the features for our regression models. We also obtained data on average temperature (in Fahrenheit) for 49 states (all except Hawaii) from 1990 to 2018 from the U.S Environmental Protection Agency. This data serves as the outcome variable for our regression models. Then, we split these two datasets into training and testing sets, using 80% of the data for training and 20% of the data for testing, which is a standard split for mid-sized datasets.

3.2 Methods

As the data initially included 21 possible features, we conducted feature selection in order to increase our models' prediction accuracy and computational efficiency. We chose to use Recursive Feature Elimination with cross validation in order to reduce the number of features because it is both easy to configure and effective at selecting features in a training dataset that are most relevant in predicting the target variable. For our first model, Linear Regression, this method only dropped one feature (population). For our second model, Random Forest, this method dropped three features (F-Gas, Waste, and Bunker Fuels), and for our third model, Support Vector Machine, this method dropped two features (F-Gas and Bunker Fuels). As none of the dropped features were related to the major greenhouse gases, we were satisfied with the output of recursive feature elimination.

Next, we conducted hyper-parameter tuning for our Random Forest regression model using Randomized Search with a standard 5-fold cross validation. We focused on three hyperparameters: bootstrap, method for sampling data points; max_depth, representing the maximum number of levels in each decision tree; max_features, representing the maximum number of features considered for splitting a node; and n_estimators, representing the number of trees in the forest. We found that the optimal hyperparameters were as follows: bootstrap = False, max_depth = None, max_features = 'sqrt', n_estimators = 400.

Without hyperparameter tuning, our Random Forest model had an average error of 1.248° F and an accuracy score of 97.48%. With hyperparameter tuning, our Random Forest model had an average error of 1.102° F and an accuracy score of 97.72%. Thus, we see that there

was a slight improvement of 0.146° F error and 0.26% accuracy. Although this difference is not large, as our initial model before hyperparameter tuning was already fairly accurate, we were satisfied with this outcome.

3.3 Results and Analysis

Using three unique regression models to predict average temperature based on greenhouse gas emissions data, we found that the Random Forest model outperformed the linear regression and the support vector machine models. As shown in Figure 1, Random Forest significantly outperformed the other two models in R^2 coefficient and mean squared error (MSE). Interestingly, the SVM model resulted in a negative R^2. This is most likely due to our model not following the trend of the data. This is further supported by the extremely large MSE for SVM.

Model	R^2	MSE
Linear Regression	0.787	15.348
Random Forest	0.971	2.111
SVM	-35.41	2,621.59

Figure 1: Using greenhouse gas emissions to predict avg temperature

Figure 2 shows comparisons between actual and predicted values of temperature for 20 random samples. There is no significant anomaly in the graph, as all predictions are of similar error from the actual values and there exists no outliers.

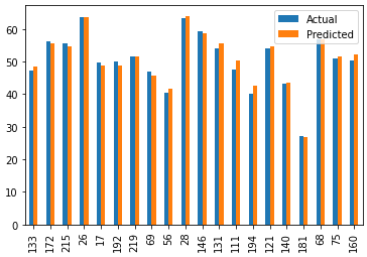

Figure 2: 20 random Samples of actual vs predicted average temperature from Random Forest Model

3.4 Discussion

Overall, our random forest model overwhelmingly outperformed both of the other models. This may be due to the fact that we performed hyperparameter tuning on this model, thus optimizing the decision trees in the forest specifically to our task. As shown in Figure 1, the SVM model resulted in a negative R^2 value, and this is most likely due to the model not following the trend of the data. One possible explanation is that the data cannot be easily linearized, resulting in the SVM model producing a fit that is not representative of trends in the data and the prediction task at hand. This does not mean that support vector machines should not be used in the future for similar prediction tasks, as they are a fairly popular model for taking climate change related prediction tasks. Instead, more kernels could be tested to find the best one that matches the task at hand.

4 Section 2: Climate Change Opinions and Political Party

4.1 Data Processing

We obtained data collected by the Yale Program on Climate Change Communication that measured and estimated the opinions and beliefs of residents living in each state and 3,142 counties with respect to key questions regarding climate change. We needed to add two features to this dataset: the percentage of the population in each of these geographic areas that are Republican and Democrat. To do so, we used data from the 2016 presidential election and calculated the percentage of the votes in each state and county that were cast for each party. This data was obtained from publicly available forums. Our goal was to build regression models that would be able to take in climate change opinion data from the Yale surveys and use the 60 estimated percentages in this data as features to predict the percentages of Republicans and Democrats for each state.

We then defined a geographic area (such as a county or state) as "Rep" (solid republican) if the percentage of republican votes exceeded the percentage of democratic votes by more than 10%. The category of "Dem" was defined in the reverse direction similarly. Meanwhile, areas whose democratic and republican vote ratios differed by less than 10% were categorized as "Tossup."

4.2 Methods and Evaluation Metrics

We considered the following regressors from the SciKit Python libraries: (1) Linear Regression, (2) Support Vector Regression, and (3) Random Forest. The linear regression model

provides a baseline for which we can determine the accuracy of the other models, and the SVM and Random Forest regressors are among the most popular regression models used to date. Additionally, with the Random Forest classifier, we optimized the following hyperparameters: number of estimators, maximum depth of each tree, and the maximum number of features per tree. We use the regressors by training the model on the climate change opinion data and political party statistics for the 3,142 counties. Then, we tested our model by allowing it to read in data on the climate change opinions of each of the 50 states in the United States and predicting the Republican and Democratic percentages for each state. As we had presidential election data from 2016 that served as our Republican and Democratic percentages for each state, our evaluation metric was the mean squared error between the predicted percentages and the actual percentages of political party affiliation. Because of the nature of the regression models, we had to build two of each regressor: one to predict the Republican affiliation and the other to predict Democratic affiliation.

In addition, for each regressor method, we performed stratified 5-fold cross validation. We maintained the same folds across each of the classifiers. We also performed feature selection by using the *f_regression* statistic in sklearn to calculate the individual effect of each feature on the regressor. This was done on the best performing overall regressor, the linear regression model. By performing feature selection on the two linear regression models (one predicting Democratic percentage and the other predicting Republican percentage), we were able to analyze the overlap between the best features in predicting Democratic percentage and the best features in predicting Republican percentage as well as determine the effects of each feature in predicting either percentage.

4.3 Data Visualization by Graphing

One of the first ways we visualized our climate change data was by creating a graph where each node represented a county. Each pair of nodes was connected with an edge if the climate change opinion data between the two nodes was sufficiently similar. To do this, we created a 60-dimensional vector for each county with each component representing one of the 60 estimated percentages in the climate change opinion dataset. Then, we calculated the norm of the difference between the vectors corresponding to each node and connected an edge if the norm was less than 10. For our 3142 counties, this edge construction gave us 27,882 edges. We also found that the largest connected subcomponent of this graph contained 2,488 of the counties, of which the majority were Republican simply because the majority of the counties in America are Republican. This initial visualization allowed us to see that the climate change opinions of the counties in America are rather similar, as evidenced by the large connected subcomponent. We extended this idea for the 50 states in America, and we found that this edge construction method yielded 67 edges and a connected subcomponent of 31 states. Because there is still a significant portion of counties and states that are not similar with regards to climate change opinion, we believe that using these 60 estimated percentages corresponding to the climate change questions would serve as viable features for our regressor models.

4.4 Results

For each of the three regressors, we predicted both the percentage of Republicans and Democrats in each state. Because of the nature of SVM and our feature selection method, we predicted these outcome variables separately and then combined them into one dataframe and calculated mean squared error from the combined dataframe. In Figure 4, we have compiled the mean squared error between the combined dataframe and our actual data for each of the regressors. In addition, for both political parties, we found the average absolute value of the discrepancy between our predicted percentage and actual percentages, shown in Figure 4.

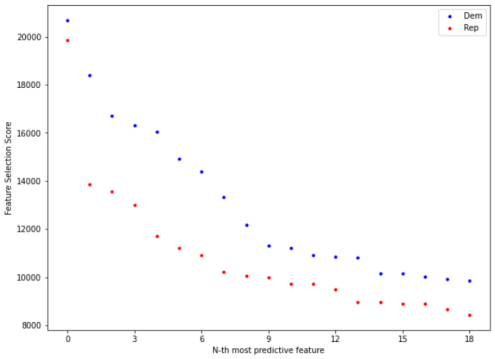

Figure 3: Feature selection score of 19 most predictive features for Democratic and Republican affiliation prediction

In addition, we also ran feature selection on our linear regression models by calculating the f_regression statistic for each of the 60 features. In general, we see from the figure above that the f_regression statistic is higher for all features when predicting Democratic percentages. Thus, we expect that our models will perform better predicting states that are solid Democratic, and this is evidenced in Figure 4 where the average percentage difference between our models and real data is smaller for the Democrat percentage for all but one of our models. When we graph the best features using the f_regression measure, we see a major drop-off in the values for this metric after around 20 features. Thus, we chose to examine the 19 best features in predicting the Democratic proportion and the 19 best features in Republican percentages. As expected, there is considerable overlap between the most telling features for the Democrat affiliation and Republican affiliation. In fact, among the 19 best features for both parties, there are 17 in common. So, our linear regression with feature selection was created by testing the model using only the 17 overlapping best features.

For our random forest model, we also performed hyperparameter tuning and found that the optimal values were n_estimators = 110, max_depth = 300, max_features = 14. This allowed us to minimize the mean squared error of the random forest model to a value of 22.832. However, even with hyperparameter tuning, we see that this model performs worse than both SVM and linear regression.

Regressor	**MSE**	**Republican % Diff.**	**Democratic % Diff.**
Linear Reg. w/o FS	16.192	3.366	2.878
Linear Reg. w/ FS	22.004	3.916	3.237
SVM	20.557	3.369	3.632
Random Forest	22.832	3.813	3.458

Figure 4: Overall MSE and Difference in Percentage by Party per Regressor

State	**MSE**	**Rank**	**Party**
North Carolina	1.943	6.5	Tossup
Florida	3.643	9.75	Tossup
Ohio	4.070	9.25	Tossup
Hawaii	4.547	11.5	Rep
Louisiana	4.587	11.75	Dem

Figure 5a: Top 5 States With Least MSE Between Regressors and Data

State	**MSE**	**Rank**	**Party**
Utah	212.532	48.0	Rep
North Dakota	181.901	47.0	Rep
New Mexico	161.963	46.75	Tossup
Maine	128.260	39.75	Tossup
West Virginia	125.543	39.5	Rep

Figure 5b: Top 5 States With Largest MSE Between Regressors and Data

We were also able to determine how each model performed in predicting the specific Republican/Democrat split for each of the 50 states. For all four models (linear regression with and without feature selection, SVM, and random forest), we ranked the 50 states based on how close our model was to predicting the correct percentages from 1 to 50, and we averaged the rank of each state across the four models in our Rank column in Figures 5a and 5b. The ranking was calculated via mean squared error, and the mean squared error column in the above tables were again averages across all 4 models for each state. Figures 5a and 5b detail the 5 states that the models were able to predict most accurately and the 5 states that the models were able to predict least accurately. As expected, the least accurate states are primarily solid Republican since the features in our climate change opinion dataset have less predictive power when it comes to Republican percentage as evidenced by our f_regression scores.

4.5 Discussion

Overall, we were able to build linear regression, SVM, and random forest models to predict the percentages of Republicans and Democrats living in a geographic area. In general, the features possess more predictive power with respect to the Democratic party affiliation statistic from their feature selection scores. This is seen by most of our regressors (with the exception of SVM) predicting the Democratic percentages more accurately. In addition, we were able to see considerable overlap in the best features for the predictions of both parties. Although the mean squared error after feature selection increased, we were still able to significantly reduce computation time at the expense of some accuracy. Finally, we were able to rank all 50 states by how well our four models were able to predict the Republican/Democrat split. In correspondence with the conclusion that Republican percentages

are more difficult to predict, the states that had the largest discrepancy between model and real-world data were primarily Republican.

5 Section 3: Predicting State Opinion Based on Emissions

5.1 Data Processing

Using the same greenhouse gas emissions (weighted by state population) data set from the U.S. EPA, we appended two additional features - average temperature per state per year, and the temperature change of each calendar year per state. This increases the total number of features to be 23. Next, from the same climate change opinion data from the Yale Program on Climate Change Communication, we extracted the percentage of each state's population that believes that climate change is really happening in 2020. This percentage will be our outcome variable for this prediction task.

Note that the feature space contains samples from 49 states and the span of 1990 to 2018 while the outcome space contains samples from 49 states in only the year of 2020. Consequently, we condensed the feature space by two methods: taking the mean for each state across the range of 29 years (we will refer to this method as by average) and taking the features from the most recent year, 2018 (we will refer to this method as by most recent year). We will investigate and compare the performance of these two methods in our prediction task. Lastly, we split the feature space and outcome space into training and test sets with a standard 80/20 split, resulting in 39 samples in the training set and 10 samples in the test set.

5.2 Methods

Using the same models we used previously to predict political party affiliation from climate change data (linear regression, random forest, and support vector machines), we additionally used greenhouse gas emissions and temperature data to predict the percentage of people that believe that climate change is happening by state.

Next, we performed feature selection using recursive feature elimination with 5-fold cross validation on each of our three models and with both of our variations of feature data - one for feature averages and one for the most recent year (2018) data. We found that the random forest model kept the most features, followed by the linear regression model, and then followed by the SVM model, which only kept 2 of the original 23 features. Some of the most frequently kept features were electric power, overall temperature change for each calendar year, energy production, and average temperature.

Then, we used randomized search with 5-fold cross validation to perform hyperparameter tuning for our random forest model. The hyperparameters that we tuned were bootstrap, max depth, max features, and the number of estimators. Additionally, we built a baseline random forest model without hyperparameter tuning to estimate the effectiveness of our hyperparameter tuning process. In comparison to the baseline model, our tuned random forest model reduced the average prediction error by 0.1435° F and increased the model accuracy by 0.22%.

5.3 Results

We built three regression models to predict the percentage of people who believe in climate change. We also tuned the hyperparameters for our random forest model, which turned out to be n_estimators = 800, max_depth = 40, and max_features = $\sqrt{n}$. After hyperparameter tuning, our random forest model was the most accurate, with the highest r^2 values as well as the least mean squared error values. In addition, we see that our random forest model performs best when considering both emissions and temperature data by the average of 29 years. Also, for this task, our SVM model performs very poorly perhaps because the features are very difficult to separate into clusters, a key indicator of models which rely on the kernel method such as SVM. These results are summarized in Figure 6.

Model	R^2 **(by means)**	R^2 **(by year)**	**MSE (by means)**	**MSE (by year)**
Linear Reg	0.1699	0.3415	8.8740	7.0392
Random Forest	0.7065	0.5920	3.1378	4.3612
SVM	-5.6130	-15.2879	70.6959	174.1230

Figure 6: Evaluation Metrics for Using greenhouse gas emissions and temperature data to predict the percentage of people who believe that climate change is happening.

In addition, we analyzed the relationship between the data and our predicted percentage of people who believe in climate change. These graphs were created by considering the models after predicting off of 29 years of data. We chose to analyze by means because this allowed our regressors to be trained off a more representative set of data with a lesser tendency to overfit off of one specific year. We see that in general predictions and actual values are fairly close and there is no specific pattern in the difference (Figures 7a, 7b, 7c). However, an interesting exception to this is our graph for the SVM regressor (Figure 7c). We see that the SVM regressor underpredicts the percentage for each of the ten test samples. This is related with the mathematical derivation of the Support Vector Machine and the effect of the margin in the prediction results of our SVM regressor.

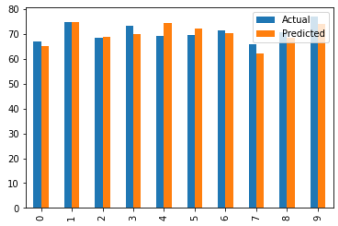

Figure 7a: actual vs predicted percentage of people who believe that climate change is happening from Linear Regression (by Means).

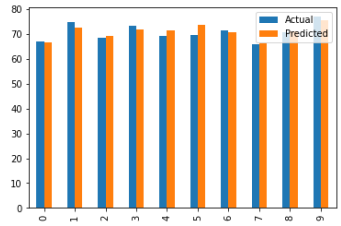

Figure 7b: actual vs predicted percentage of people who believe that climate change is happening from Random Forest (by Means).

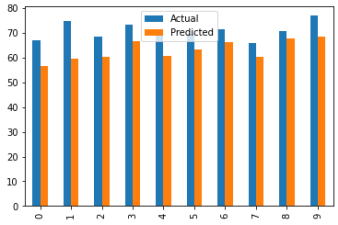

Figure 7c: actual vs predicted percentage of people who believe that climate change is happening from Support Vector Machine Regressor (by Means).

5.4 Discussion

Some of the most frequently kept features were electric power, overall temperature change for each calendar year, energy production, and average temperature. It is interesting to note that these top features, two of which relate to temperature and two of which relate to the manner in which we produce energy, were relatively indicative of the percentage people of a certain state believe in climate change. The continuous push towards alternative energy could provide a dramatic - most likely beneficial - shift in this relationship.
After building three models to predict the percentage of climate change believers based on greenhouse gas emissions and temperature data, we found that the random forest model outperformed the linear regression and support vector machine models. One limitation we faced was that we did not find enough data for climate change opinion. Our feature space of merged emissions and temperature data includes data for 49 states from the years 1990 to 2018 while we only had 2020 data of the percentage of people that believe that climate change is happening for 49 states. In order to combat this issue, we condensed the feature space with two methods: taking the average of each feature and taking features only from the most recent year (2018). One consequence of this solution is that we produced small train and test sets, and predicting on these small data sets can produce volatile results from our models.

6 Spotlight Method: Random Forest

In order to understand the Random Forest Regressor, we must first understand decision trees, also known as decision trees. Decision trees are constructed by recursively diving the input space into different regions with nodes that partition the data set. Each decision tree is fit to a subset of the training data that are randomly selected with replacement. In order to find the best partitioning of data, we use a greedy procedure to compute a locally optimal MLE. This procedure determines if the gain of using this feature is large enough to justify the creation of a new node in the decision tree, thus ensuring that a single feature does not hold too much weight in the regression model.

Decision trees are derived from the assumption that it is difficult to construct a good kernel function that compares the similarity between data vectors. So, decision trees try to extract useful features directly from the dataset itself. To do so, they incorporate an adaptive basis-function model (ABM) in the form of

$$f(x) = w_0 + \sum_{m=1}^{M} w_m \phi_m(x) \tag{1}$$

There are many advantages to using decision trees. They are very intuitive to interpret, malleable, insensitive to monotone transformations of inputs, and are robust to outliers. However, they are not very accurate because of the greedy nature of the algorithm by which these trees are constructed. Because small changes to input data can cascade into large variations in the tree structure, decision trees are known as high variance estimators. It is with this drawback of decision trees in mind that we explore the random forest classifier.

Random Forest is an ensemble learning classifier based on having K decision trees. Each decision tree arrives at its own prediction outcome. The random forest model then returns the outcome variable with the most votes from its K decision trees. Having multiple trees increases the accuracy of the classifier and reduces the likelihood of the model to overfit the data. This is because while each individual decision tree may overfit the training data, this method combines multiple weak hypotheses from these decision trees into one strong hypothesis from the model.

Random Forest minimizes the cost of using decision trees by training many different trees on subsets of the input data and aggregating the results in a technique known as bagging. Bagging runs the same classifier on different subsets of data, leading to the construction of very highly correlated predictors. In order to circumvent the high variance that this would cause, Random Forest decorrelates the base learning algorithms by limiting the trees to a random subset of input variables in addition to a random subset of the training data.

To consider factors such as the size of the random subset of input variables, we performed hyper-parameter tuning using K-fold cross validation, ensuring that each decision tree contains enough depth to capture relevant information while not promoting too much underfitting. We first optimized the fold accuracy by the number of estimators, which is simply the number of trees in the forest. Then, we considered the maximum depth of the trees. This variable is extremely important because it is directly related to the cost of adding a new split to the trees, which affects how the trees and leaves are formed. Finally, we tuned the number of features to consider when looking for the best split (max features).

7 Conclusion

All in all, this project focused on how we could use machine learning to explore the sociopolitical side of climate change. Initially, we used each state's greenhouse emissions data to predict the state's temperature. Our random forest model performed best for this task. Next, we evaluated how climate change opinions within states and counties could be utilized to predict the political party affiliation of citizens in these states. We determined that Democratic percentages were more easily predicted than Republican percentages, and our linear regression model performed best. Finally, we used each state's emissions and temperature data to predict climate change opinion. Again, random forest after hyper-parameter tuning outperformed both linear regression and SVM. From this work, we show the intricate connections between the general American public's opinion regarding climate change and a myriad of other scientific data such as greenhouse emissions, temperature, and political party affiliation. Future work could explore why Democratic percentages were easier to predict, as well as using different kernel methods in SVM to better understand the relationship between emissions and temperature data with climate change opinion. We believe that the questions that our tasks raised are important to explore for scientists, educators, and policymakers to better understand the interconnected nature of climate change science, policy, and social perception. Until we fully understand these issues, we will continue to struggle move forward with more environmentally friendly policies to reduce our carbon footprint.

References

[1] Briant, Elliot. "Understanding Climate Change with Machine Learning." Medium, MyTake, 24 Oct. 2019, medium.com/mytake/understanding-climate-change-with-machine-learning-fb45a047dd2b.

[2] Howe, P., et al. "Geographic Variation in Opinions on Climate Change at State and Local Scales in the USA." Nature Climate Change, vol. 5, 6 Apr. 2015, pp. 596–603., doi:https://doi.org/10.1038/nclimate2583.

[3] "Climate Watch - U.S. States Greenhouse Gas Emissions ." Climate Watch, World Resources Institute, 2021, datasets.wri.org/dataset/climate-watch-states-greenhouse-gas-emissions.

[4] "Comparative Climatic Data." National Centers For Environmental Information, National Oceanic and Atmospheric Administration, 2018, www.ncdc.noaa.gov/ghcn/comparative-climatic-data.

[5] "The Politics of Climate Change in the United States." Pew Research Center Science; Society, 4 Oct. 2016, www.pewresearch.org/science/2016/10/04/the-politics-of-climate/.

[6] Marlon, Jennifer, et al. "Yale Climate Opinion Maps 2020." Yale Program on Climate Change Communication, Yale College, 2 Sept. 2020, climatecommunication.yale.edu/visualizations-data/ycom-us/?fbclid=IwAR0odcD8gUbVc7uu8uUgKiyjl9ZZei-NFYF38Wz-lXXnFlx0h63iKwq8u5Y.

[7] "The Politics of Climate Change in the United States." Pew Research Center Science; Society, 4 Oct. 2016, www.pewresearch.org/science/2016/10/04/the-politics-of-climate/.

[8] Zheng, Harvey. "Analysis of Global Warming Using Machine Learning." Computational Water, Energy, and Environmental Engineering, vol. 07, no. 03, July 2018, pp. 127–141., doi:10.4236/cweee.2018.73009.

8 Appendix

Summary of each member's contribution to the project: Sara focused on the section 1 code and write-up, Brian focused on the section 2 code and write-up, and Darren focused on the section 3 code and write-up. However, all three of us worked synchronously to research our topic, find suitable databases, write the code, make the poster, debug, and write up our work. More specifically, Sara was most confident in finding, creating, and handling dataset processing; Brian was most confident in handling graph analysis; and Darren was most confident in feature selection and building regression models. Thus, we all communicated with each other to lend a helping hand in each other's sections, especially when group members hit a road block with errors or unusual behavior in their code. Overall, each group member contributed an equal amount and time and work in this assignment.