University of Leeds

School of Computing

COMP5122M – Data Science

Coursework: Global Climate Attitudes and Eco-Anxiety — Part II

Team Name: Meowth

Kaustubh Kaushal [wftn0343]
Siddhant Ram [fmtm0110]
Ziyun Kang [csdf0983]
Qianning Zhou [cxvk0212]

1. Introduction and Aim

Through the analysis of demographic factors such as country, age group, and education, as well as a set of attitudes related to climate change, this report investigates public support worldwide for bolstering national climate commitments. The study uses predictive modeling with data from The People's Climate Vote 2024[1] (UNDP, 2024) to determine whether support for stronger climate pledges is more influenced by attitudinal or demographic factors. In order to understand how countries can be categorized based on public perceptions and concerns about climate change, a second goal is to perform a clustering analysis to group countries based on similarities in climate-related attitudes.

The predictive modeling component contrasts Random Forest and Logistic Regression, two supervised learning models. K-Means is used in the clustering component to aggregate attitude variables at the Country level. By preprocessing the data, defining the target, training models, comparing performance, and interpreting feature contributions, the report employs an organized empirical methodology. Cross-regional comparison of national attitude profiles is further supported by the clustering.

2. Data Overview and Preprocessing

Weighted national response percentages to survey questions about climate change make up the dataset. Country, age group (Under 18, 18–35, 36–59, 60+, All Ages), education level (Never attended, Under 12, 12–19, 20+, All Education), and several attitude questions about climate awareness, concern, and behavioral influence are used to segment responses.

Predictive Modelling Preprocessing

- Target variable: Support for strengthening climate commitments was coded as a binary variable:
  y=1 if "Strengthen", otherwise y=0.
- Features included:
  - Demographics: Country, Age, Education
  - Attitudes: 10 selected variables:

Attitude	Response
"Worry for next generation"	"Extremely", "Very", "Somewhat
"Extreme weather experience"	"Worse than usual"
"Thinking about climate change"	"Daily", "Weekly"
"Effects on big decisions"	"A lot"
"Renewable energy transition"	"Very quickly", "Somewhat quickly"
"Protect and restore nature"	"A lot"
"Country performance"	"Very well", "Somewhat well"
"Big businesses performance"	"Very well", "Somewhat well"
"Teaching about climate change"	"More", "More education"
"Rich countries helping poor"	"More help"

- Merging: Target rows were merged with attitudes using country, age, and education.
- Missing values were imputed (most frequent for categorical, mean for numeric), and numerical attitudes were standardised.
- The global summary row was removed to avoid duplication and bias.

Clustering Preprocessing

To maintain consistency, clustering analysis was limited to Country × All Ages × All Education. Before performing K-Means, StandardScaler was applied and the ten attitude variables were kept. To choose a suitable cluster structure, silhouette scores were calculated for K = 2–7.

3. Methodology

Predictive Modelling

Two models were chosen to reflect different modelling philosophies:

1. Logistic Regression – linear separation, interpretable coefficients.
2. Random Forest – non-linear ensemble method, feature importance extraction.

(Separate dataset to the train set and test set as 70% and 30%.)

Both models were evaluated using:

- Accuracy[2]
- F1 Score[2]
- Balanced Accuracy[2]
- ROC-AUC[2]
- Confusion matrices
- Feature importance / coefficient plots

Clustering Analysis

- Selecting attitude features and constructing a country-level attitude matrix.
- K-Means clustering, using silhouette scores to determine the best K.
- Cross-regional comparison by mapping countries to world regions to examine whether clusters reflect geographic patterns.

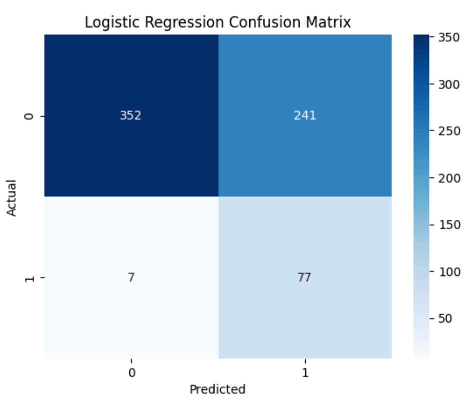
Cluster profiles were interpreted using the mean values of the 10 attitude measures per cluster.

4. Predictive Modelling Results

4.1 Model Performance Comparison

Logistic Regression

- Accuracy: 0.633
- F1 Score: 0.383
- Balanced Accuracy: 0.755
- ROC-AUC: 0.717



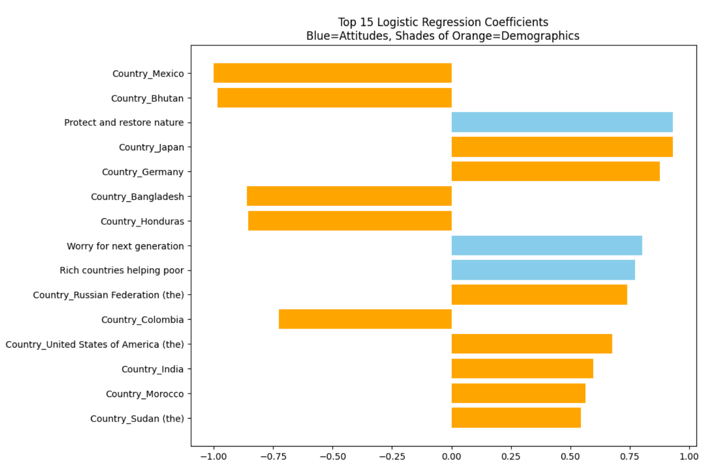
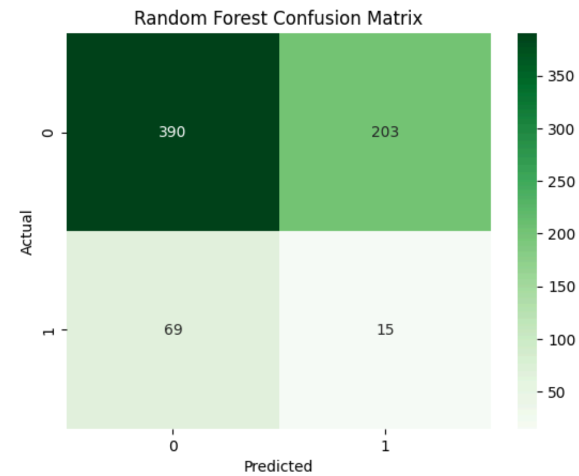


Figure 1 and 2. Logistic Regression Confusion Matrix and Logistic Regression Top 15 Coefficients (Demographics vs Attitudes)

Random Forest

- Accuracy: 0.598
- F1 Score: 0.099
- Balanced Accuracy: 0.418
- ROC-AUC: 0.619



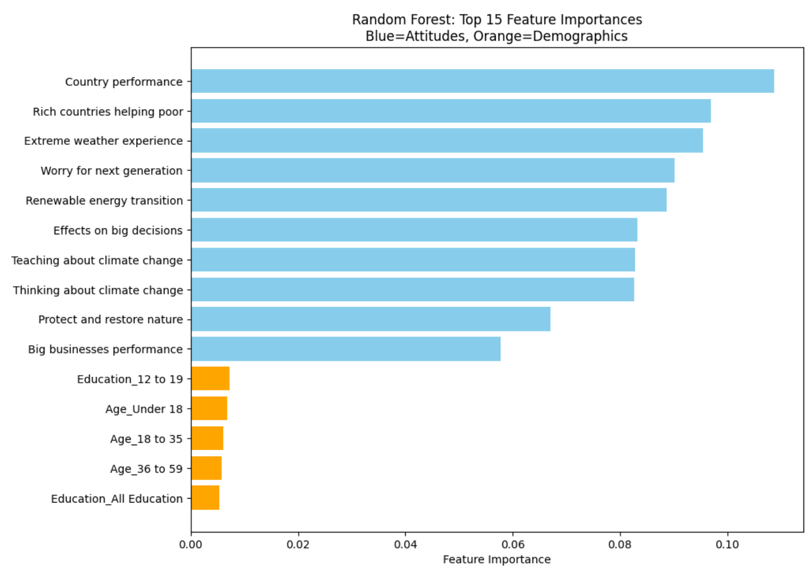


Figure 3 & 4. Random Forest Confusion Matrix and Random Forest Top 15 Coefficients (Demographics vs Attitudes)

In every metric, logistic regression performs better. The Random Forest had trouble identifying the minority class (countries/segments not supporting strengthening), as evidenced by its low F1 and balanced accuracy despite threshold tuning (best threshold = 0.10). The ROC-AUC of 0.717 for logistic regression indicates a moderate level of predictive power, but it is significantly higher than Random Forest.

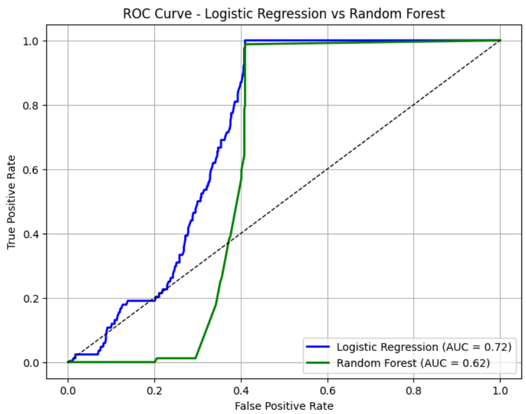


Figure 5. ROC Curve Comparing Logistic Regression and Random Forest

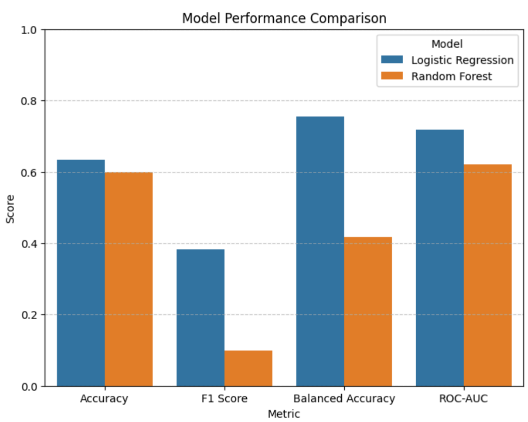


Figure 6. Comparison of Model Performance Metrics (Accuracy, F1 Score, Balanced Accuracy, ROC-AUC)

4.2 Which Matters More: Demographics or Attitudes?

These two models demonstrate that different patterns of support for bolstering climate commitments are influenced by demographics or by attitudes. Random Forest result shows that attitudes factors are dominant in Figure 4 as values of demographic variables are quite low and Figure 2 illustrates that demographic factors, particularly country account for the majority of feature importance in the Logistic Regression, and attitude coefficients, “Protect and restore nature”,“Worry for next generation” and “Rich countries helping poor” have a big impact on strengthening commitment as well. Demographic variables are one-hot encoded and the coefficients of it accumulate into a large total while attitude variables are fewer or scaled differently and it displays as smaller numerical coefficients. Hence, these graphs still imply that the main causes of variation in support are still structural demographic differences rather than personal perceptions of the climate.

4.3 Interpretation

Logistic regression outperformed Random Forest in terms of accuracy, F1 score, balanced accuracy, and ROC-AUC. While the Random Forest had trouble with the sparse one-hot encoded demographic features, this suggests more pronounced linear relationships in the data. The predominance of demographic predictors in both models indicates that support for more robust climate commitments varies more by age group and country than by attitudes. These results are consistent with more general worldwide trends in which national context has a greater influence on climate attitudes than individual climate concern.

5. Clustering Analysis

5.1 Selecting K

Silhouette scores ranged from 0.1712 to 0.3075. Although values are generally low due to the similarity of national averages, the best score was at **k = 2**, but we selected **k = 3** because it achieved a high silhouette score while producing the most interpretable and policy-relevant cluster structure. k = 2 collapsed important distinctions, while k > 3 fragmented the data into less meaningful subgroups.

This indicates that global attitudes form several small but meaningful clusters, rather than a simple split.

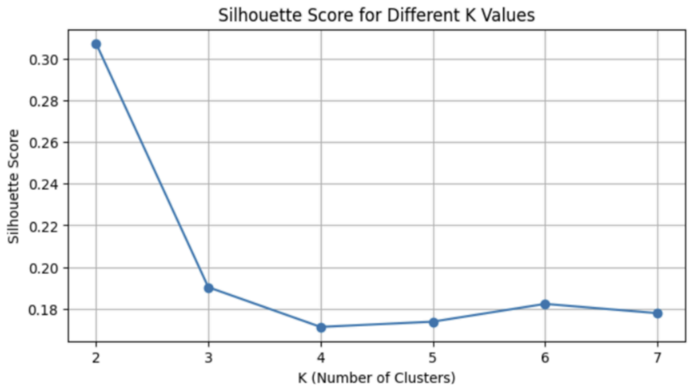


Figure 7.Silhouette Scores for K = 2 to 7

5.2 Cluster Profiles

Cluster means show minimal differences across the ten attitude measures. All values fall within a narrow range of roughly 17–31%, with variations between clusters typically below one percentage point. This indicates that global climate attitudes are highly consistent and differ only slightly between countries. Such limited variation explains why attitudes contribute less to prediction compared to demographic factors and supports the conclusion that demographic context is more influential than attitudinal differences, because attitudes vary little, the model has limited signal to separate countries, leading to overlapping clusters.

5.3 Regional Distribution of Clusters

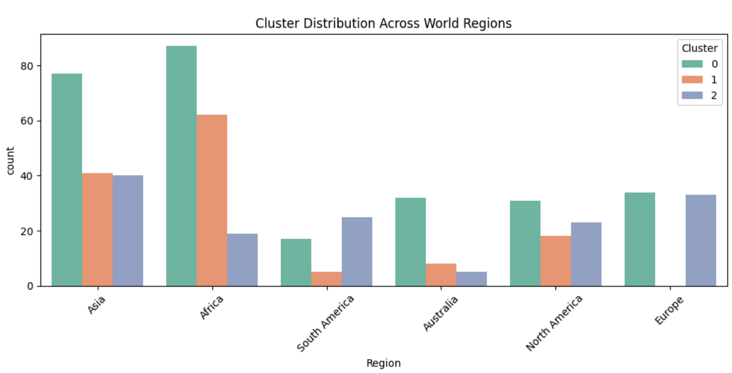


Figure 8. Cluster Distribution Across World Regions

The cross-regional mapping shows:

- Africa: spread across Clusters 0, 1, 2
- Asia: distributed across all clusters
- Europe: Mainly Clusters 0 and 1
- North America: present in all three clusters
- Oceania & South America: relatively even distribution

Cluster dispersion indicates that shared climate attitudes cross regional boundaries, reinforcing the idea that climate concern is a global phenomenon rather than region-specific. North America is spread across three clusters, Europe is primarily found in Clusters 0 and 1, and Africa and Asia are present in every cluster. Mixed representation is also seen in South America and Oceania. This dispersion shows that there is no geographical concentration of nations with comparable attitude profiles. Rather, there are minor variations in attitudes within each region, which supports the idea that attitudes around the world are generally uniform and not significantly influenced by geography.

6. Conclusion

- The predictive analysis shows that demographic factors, especially country and age, are stronger determinants of support for strengthening climate commitments than attitudinal variables.
- Logistic Regression performed better than Random Forest, indicating that the relationships in the data are mostly linear and that Random Forest struggled with sparse one-hot encoded demographic features.
- Attitude variables contributed less to prediction, as they showed very limited variation globally, typically ranging only between 17%–31%.
- The clustering analysis confirmed this consistency: countries grouped into clusters with only minor differences, and no clear regional patterns emerged.
- These findings suggest that public attitudes toward climate change are globally stable, while differences in support are driven mainly by demographic and national context.
- Policymakers should therefore tailor communication strategies by demographic segments and consider country-specific contexts when promoting stronger climate pledges.

Appendix

Appendix A: Predictive Modelling

Step Category	**Description**
Target variable	Support for strengthening climate commitments was coded as a binary variable:y=1 if "Strengthen", otherwise y=0.
Features	• Demographics: Country, Age, Education • Survey Attitudes: 10 selected variables
Model Performance Comparison: Logistic Regression vs. Random Forest	Logistic Regression: Accuracy: 0.633, F1 Score: 0.383, Balanced Accuracy: 0.755, ROC-AUC: 0.717
	Random Forest: Accuracy: 0.598, F1 Score: 0.099, Balanced Accuracy: 0.418, ROC-AUC: 0.619
Demographics or Attitudes matter more?	Climate commitments are more strongly influenced by demographics than by attitudes.

Appendix B: Clustering Analysis

Step Category	**Description**
Clustering Preprocessing	StandardScaler was applied and the ten attitude variables were kept.
Selecting K	Using silhouette scores to determine the best K. Best K = 3
Cluster Profiles	Cluster means fall within a narrow range of roughly 17%–31%
Compare clusters across world regions	Cross-regional comparison by mapping countries to world regions to examine whether clusters reflect geographic patterns.

References

[1] United Nations Development Programme (UNDP).2024.*The Peoples' Climate Vote 2024.*[Online]. [Accessed 9 December 2025]. Available from: https://peoplesclimate.vote/data-center
[2] Metrics and scoring: quantifying the quality of predictions .[Online]. [Accessed 11 December 2025]. Available from: https://scikit-learn.org/stable/modules/model_evaluation.html