Examining the UN General Assembly Debates to Predict Peace

Abhinav Agnihotry 15667766, Brent Brakenhoff 13068989,
Jonathan van der Maas 15248844, Shantanu Motiani 13757830

Group A3

1 Abstract

As global challenges such as conflict, inequality, and political discourse continue to shape the world, understanding the factors that contribute to peace and quality of life becomes increasingly crucial. This projects aims to examine the United Nations General Debate Corpus to see what these speeches reveal about a country's history of conflict and what this says about the quality of life of its inhabitants. Which poses the question: To what extent does a country's General Debate speech and Happiness Index correlate with its Global Peace Index? This report concludes that there is a correlation between the sentiment score of a country's speech and that country's speech score, this correlation was made more significant by taking the rolling average of the sentiment score.
Additionally this project will attempt to predict a country's the Overall Peace Index based on its General Debate speech and its Happiness index, which poses the question: How accurately can a model predict a country's Global Peace Index based on the content of its General Debate speech and its Happiness Index? A random forest regression model is able to predict a country's overall peace index fairly well. Based on the sentiment score of that country's speech, that country's Happiness index and its ongoing conflict score, it achieves an R^2 of 0.874 with an MSE of 0.027 compared to a Linear Regression model which achieves an R^2 of 0.806 with an MSE of 0.068.

2 Introduction

2.1 Background

The theme of this year's General Assembly of the United Nations is "Leaving no one behind: acting together for the advancement of peace, sustainable development and human dignity for present and future generations" (*United Nations General Assembly*, 2024), underscores the importance of peace, aligning with the Sustainable Development Goal (SDG) of "Peace, Justice, and Strong Institutions" (*SGD 16: "Peace, Justice, and Strong Institutions"*, 2024). Each year, the General Assembly opens with a General Debate session where countries present statements reflecting their concerns and priorities. These statements serve as indicators of each country's political landscape, and analyzing them can yield valuable insights into current events. To facilitate this analysis, the United Nations General Debate Corpus was created, encompassing statements from 1946 to 2023 (Jankin, Baturo, & Dasandi, 2017).

In light of recent international conflicts, this project aims to analyze General Debate statements to identify correlations between a country's statements and its sociopolitical landscape. Since the corpus includes only the debate statements, additional datasets are required to represent the sociopolitical state of each country. To measure peace, the Global Peace Index, which ranks countries based on 23 metrics related to internal and international peace, was selected (*Global Peace Index 2008-2023*, 2023).
To gain a better perspective on the social state of each country we decided to use the World Happiness index, which contains a score for each country based on the happiness of its inhabitants. This score is computed based on a multitude of different variables, including as GDP per capita, life expectancy and perceptions of corruption (Helliwell et al., 2024).

2.2 Research Questions

This report seeks to investigate the relationship between the content of countries' General Debate speeches at the United Nations, the Happiness Index and their broader sociopolitical and well-being indicators, specifically focusing on the Global Peace Index. By examining these elements, the analysis is guided by two key research questions:

1. **To what extent does a country's General Debate speech and Happiness Index correlate with its Global Peace Index?** This question seeks investigates the correlation between a nation's UN General Debate speech content, its sentiment, Happiness Index, and Global Peace Index, aiming to determine if patterns are indicative of a country's peace status.

   Our hypothesis for the first research question is that the sentiment expressed in a country's UN General Debate speech, along with its Happiness Index, is positively correlated with its Global Peace Index. This expectation is based on the belief that well-being and positive communication in international forums reflect more stable and peaceful conditions within a country.

2. **How accurately can a model predict a country's Global Peace Index based on the content of its General Debate speech and its Happiness Index?** This question evaluates the predictive capability of models in forecasting a nation's Global Peace Index, exploring whether patterns within the features can accurately anticipate a country's levels of conflict, safety, and security.

   Our hypothesis for this research question is that a machine learning model can accurately predict a country's Global Peace Index using data from its General Debate speech and Happiness Index. This is based on the assumption that both a country's internal well-being and the tone of its international discourse are strong indicators of its overall level of peace and security.

3 Methodology

3.1 Data Preprocessing

As mentioned in the introduction, the main dataset used for this project is the United Nations General Debate Corpus 1946-2023. For every year this dataset consists of the 'session', the year the current session was held; the 'ISO-alpha3 Code', which is an abbreviated version of the country name; and the speech itself. Since the data was provided in .txt format, we utilized Python's *os* library to read the files and consolidate them into a single *Pandas* DataFrame. This process also involved extracting and retaining the country code and year information from the file paths of the individual documents to ensure accurate merging. After merging the speeches, Python's NLTK library was employed to preprocess the speeches

- **Text cleaning**: The raw speech text contained noise such as URLs, numbers, non-alphabetic characters, and common stopwords (e.g., "the", "and"). These were removed using regex-based text cleaning techniques.
- **Ignoring common UN words**: To focus on the more meaningful words, frequent and generic words often used in UN contexts (such as "nations", "united", "international", etc.) were also excluded during text processing.
- **Sentiment analysis**: The sentiment polarity score of each speech was calculated using the `nltk` library's `VADER` sentiment analyzer, allowing us to numerically quantify the sentiment conveyed by the speeches on a scale ranging from -1 to 1.

The Global Peace Index dataset consists of 4 different variables for every country for every year. The 'Overall Scores', depicting the overall peace score of a country with lower meaning a country is more peaceful; 'Safety and Security', depicting the degree of safety and security in that country; and 'Ongoing Conflict', which measures if a country is currently facing national or international conflict.
The World Happiness Index includes several variables, with the main one being Life Ladder, which reflects the general happiness score for each country. The other variables are either self-evident or less relevant to this research.
These three datasets was merged by country name and year, creating a comprehensive dataset for analysis. As both the Peace and Happiness indices start in 2008, the analysis does not include data from prior years. The peace index was inverted so that higher scores indicate greater peace.

3.2 Exploratory Data Analysis

The EDA process aims to explore the relationship between the speeches given by countries at the United Nations General Assembly (UNGA) and their peace index scores over the years. The following steps outline the EDA methodology, leading up to the main question: *Is there a correlation between the sentiment of a country's speech and its peace index?"*.

3.2.1 EDA stages

The following questions were tackled in a step-wise order to arrive at the final results of the EDA:

1. **How does the Peace index (Overall Scores) correlate with a country's Happiness index (Life Ladder)?**
    - A plot of the means of these two variables over the years was created to have a visual clue of the linear correlation.
    - The Pearson correlation between these two variables was taken to examine the degree of linear correlation.

2. **Is there a correlation between the peace index and peace (or conflict) related words in the speech?**
    - The pairwise correlation between the peace index and various factors, such as social support, healthy life expectancy, and perceptions of corruption, was calculated.
    - A heatmap was generated to visually display the correlations.
3. **Is there a correlation between the sentiment of a country's speech and its peace index?**
    - The sentiment polarity of each speech was calculated using the `VADER` sentiment analysis tool from the `nltk` library.
    - A scatter plot was created to show the relationship between the speech sentiment score and the peace index for each country.

Since the graphs containing the peace index and sentiment score for separate countries appeared to show correlation but not too much because of the volatile nature of the sentiment score, the next step was to compare the peace index with a rolling average of the sentiment score to reduce the volatility of the variable. This showed a clear correlation between the sentiment score and the peace index.

3.3 Prediction

To predict the overall peace index using socio-economic and conflict-related predictors, several regression models were evaluated, including Linear Regression, Ridge and Lasso Regression, and Random Forest Regressor. Each model was assessed based on its ability to minimize prediction error and accommodate the data's complexities.

The dataset consisted of various predictors, including rolling sentiment scores, socio-economic indicators, and conflict metrics. Initial preprocessing involved handling missing values by excluding incomplete rows. Correlation analysis helped identify an initial set of predictors, which was refined using adjusted R^2 calculations and pair plots to check for collinearity. The final selected predictors were:

- **Sentiment Score**: A rolling average of sentiment scores of the different UN General Assembly speeches between 2008 and 2023.
- **Life Ladder**: A well-being measure from the World Happiness Report.
- **Ongoing Conflict**: A metric for the extent of ongoing domestic and international conflict related to each country.

The modeling process began with a Linear Regression to establish a baseline. This model assumes a linear relationship between the predictors and the target variable. Residual plots revealed issues because of the variance of residuals increasing with the predicted values. This pattern suggested that the linear model was unable to fully capture the complexity of the data, prompting further exploration of regularized models.

Next, Ridge Regression was employed to address potential collinearity and overfitting. Ridge regression applies an L2 regularization term, shrinking the coefficients of less relevant features while still retaining all predictors. Lasso Regression was explored as an alternative regularization method. Lasso applies L1 regularization, which can shrink some coefficients to zero, effectively performing feature selection. Despite the improvements offered by Ridge and Lasso, the presence of non-linear relationships between the predictors and the target variable, as evidenced by the residual analysis, suggested the need for a more flexible model. Therefore, a Random Forest Regressor was introduced. Random Forests, an ensemble learning method, combine multiple decision trees, each trained on bootstrapped samples of the data, to capture both linear and non-linear interactions. This method also averages the predictions of the individual trees, reducing the risk of overfitting while improving predictive performance.

The Random Forest Regressor was trained using 100 decision trees (`n_estimators=100`), with a random seed (`random_state=42`) for reproducibility. Ultimately, The Random Forest Regressor proved to be the most suitable model for this task. The Random Forest's ability to model non-linear relationships allowed it to capture non linear interactions between the predictors. Furthermore, the ensemble nature of the random forest reduced the risk of overfitting by averaging the predictions from multiple decision trees, leading to improved model robustness. Therefore the Random Forest Regressor was chosen as the final model.

4 Results and Discussion

4.1 EDA

Figure 1 shows that there does seem to be a correlation between the Overall Peace score and the Life Ladder score, although it is not a strong correlation. These two variables have a Pearson Correlation of 0.49, which supports the assumption that these metrics are positively correlated, although again not strongly.

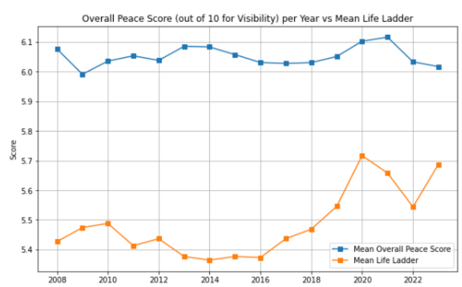

Figure 1: Mean of peace index and happiness index over the years.

This lead to the question of if the sentiment scores of each country's speeches over time can show correlations with the peace index. Firstly, we plotted a general scatterplot which can be seen in Fig 2:

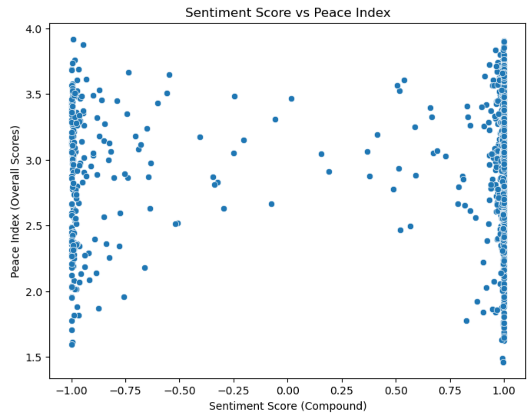

Figure 2: Scatterplot of Peace Index against Sentiment score

The scatterplot in Fig 2 shows no specific trend to conclude that the sentiment of UN speeches have a clear correlation with the peace index of that country. Moreover, we see very little variation in the sentiment scores since most of them are concentrated extremely close to -1 or 1. To dive deeper into this, we wanted to check how often do countries with different peace index scores mention peace-related terms (e.g., "peace," "conflict," "war") and conflict related terms (e.g., "war", "violence", "terrorism", "casualities") in their UNGA speeches.

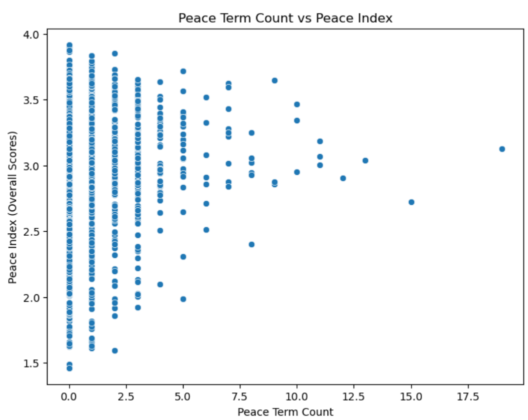

Figure 3: Scatterplot of peace related terms in the speech and the peace index

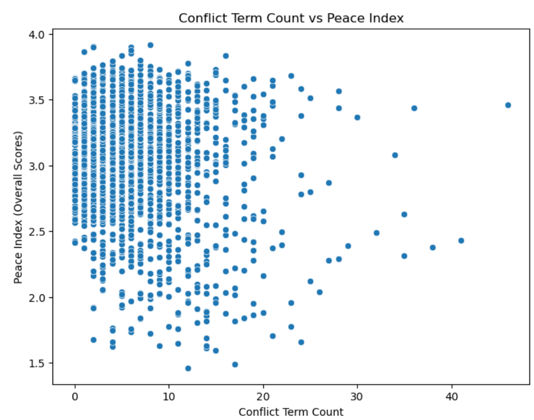

Figure 4: Scatterplot of conflict related terms in the speech and the peace index

The Figure 3 & 4 above show a slight positive correlation for peace terms and a slightly negative correlation for conflict terms which is not strong evidence but leads to conclusion that there is some predictive power in the sentiment scores to predict peace. This can be seen in specific examples of countries whose peace index and UN speech sentiment scores were drastically affected by real world events in the figures 5 & 6.

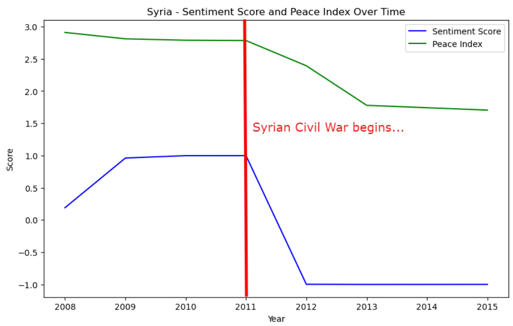

Figure 5: Impact of civil war on Syria's peace index and speech sentiment

Figure 6: Impact of Euromaidan protests and Russian Invasion

4.2 Prediction

To predict the overall peace index, several regression models were tested, including linear regression, ridge regression, lasso regression, and the random forest regressor. The performance of each model was assessed based on their ability to generalize to unseen data and handle potential issues such as collinearity and non-linear relationships between the predictors and the target variable.

Initially, the linear regression model was applied, which provided a reasonable baseline performance, achieving an R^2 score of 0.806 on the test set. However, analysis of the residuals revealed increasing variance as predicted values increased (see Figure **??**). This pattern suggested that the linear model struggled to fully capture the complexity of the data, indicating that more advanced methods could improve performance.

To address potential collinearity and overfitting, Ridge regression was applied. After hyperparameter tuning with GridSearchCV, the optimal model with $\alpha = 1$ also achieved an R^2 score of 0.806 on the test set. Ridge regularization helped stabilize the model by shrinking coefficients without eliminating any features, which made sense given that feature selection had already been performed before model training.

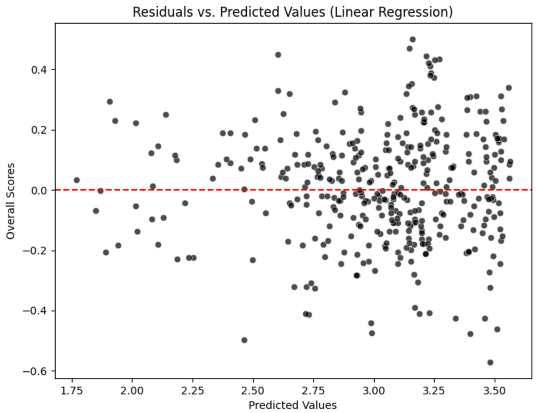

Figure 7: The residual plot of the initial linear regression model

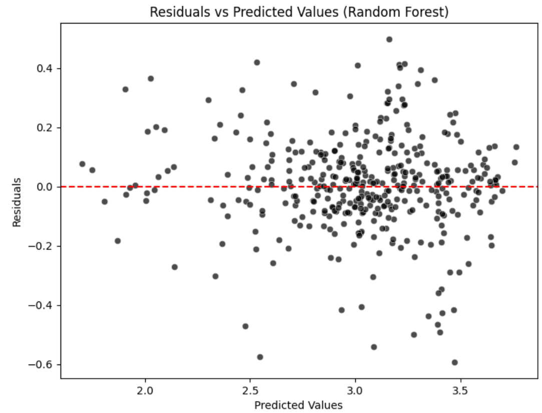

Figure 8: The residual plot of the final random forest model

In parallel, Lasso regression was explored as an alternative regularization method. However, after tuning the regularization parameter, the best lasso model with $\alpha = 0.01$ achieved a slightly lower R^2 score of 0.798. The performance of Lasso was similar to Ridge, but its reduced ability to improve performance suggested that further feature selection did not add significant value. This aligns with the fact that an initial feature selection step was conducted prior to training.

Finally, a random forest regression was introduced to capture non-linear relationships in the data. After training the model with 100 decision trees, the random forest achieved an R^2 score of 0.847 improving on the earlier linear models. Apart from the increased performance, Figure 8 also shows that the random forest captures non-linear interactions between predictors as the increasing variance with predicted values that was observed in Figure 7 is not seen in the residual plot of the random forest model making it the most suitable choice for predicting the peace index.

Model	R^2	MSE
Linear Regression	0.806	0.068
Lasso	0.798	
Ridge	0.806	0.034
Random Forest	0.847	0.027

Table 1: This R^2 and MSE comparison of the predictive models.

The predictive models developed in this study aimed to explore the relationship between the sentiment of a country's speech at the UN General Assembly, the happiness within that country, a measure of domestic and international conflict the country is involved in and its peace index. Several models were evaluated, with the random forest regressor outperforming linear regression, Ridge, and Lasso models. The ridge and lasso models, which introduced regularization to address possible collinearity issues, offered improvements over the basic linear regression model. However, neither model could outperform the random forest in terms of mean squared error or its ability to handle non-linearities and interactions between features.(see Table 1)

4.3 Limitations

The key limitation of the model is the relatively small amount of data used, covering the period from 2008 to 2023. This limited dataset constrains the model's ability to capture longer-term trends and relationships between the variables. The impact of this limitation is particularly significant for the five-year rolling average applied to the sentiment scores of the UN speeches, as it reduces the amount of effective data available for modeling. The rolling average smooths fluctuations in sentiment but also relies on a sufficient historical span to provide meaningful insights. A very interesting direction for future research would be to further look into countries with a very strong negative correlation between the peace index and the rolling average of the speech sentiment. Further research into which topics they discuss during the general assembly could show what is most important for countries who are, for example, rebuilding after a war. This would not only help policy makers but also aid organisations. A larger dataset incorporating more historical data could be used to train time series models, enabling analysis of country-specific trends and improving predictive accuracy. This approach would also provide deeper insights into the impact of features within the context of each country's unique historical trajectory.

5 Conclusion

This study explored the relationship between the sentiment of United Nations General Assembly (UNGA) speeches, a country's happiness index, and its Global Peace Index (GPI). The findings highlight that while sentiment analysis of UNGA speeches offers some correlation with peace indices, the relationship is not straightforward. The rolling average of sentiment scores proved more effective in capturing meaningful correlations, particularly in cases where real-world events—such as civil wars or political upheavals—directly impacted both speech content and peace index scores. Notably, countries experiencing conflict, such as Syria and Ukraine, exhibited distinct patterns in sentiment and peace index fluctuations, suggesting that speech sentiment is reflective of a country's socio-political climate to some extent.

From a predictive standpoint, machine learning models were employed to forecast a country's peace index based on speech sentiment, happiness scores, and conflict metrics. Of the models tested, the Random Forest Regressor demonstrated superior performance, achieving an R^2 score of 0.847 and a mean squared error (MSE) of 0.027, outperforming traditional linear and regularized regression models. This success is attributed to the Random Forest's ability to capture non-linear relationships and interactions between the predictors, making it the most suitable model for predicting GPI in this context.

Despite these findings, the study is constrained by the limited time span of the data (2008–2023) and the volatility of speech sentiment scores, which impacted the robustness of the analysis. Future research could benefit from a more extended dataset to explore long-term trends and incorporate time series modeling. Additionally, further investigation into speech content, particularly for countries with strong deviations in sentiment and peace index, could offer valuable insights for policymakers and aid organizations focusing on peacebuilding and conflict resolution.

References

Global peace index 2008-2023. (2023). Institute for Economics and Peace. Retrieved from http://visionofhumanity.org/resources

Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B., & Wang, S. (Eds.). (2024). *World happiness report 2024.* Wellbeing Research Centre, University of Oxford.

Jankin, S., Baturo, A., & Dasandi, N. (2017). *United Nations General Debate Corpus 1946-2023.* Harvard Dataverse. Retrieved from https://doi.org/10.7910/DVN/0TJX8Y doi: 10.7910/DVN/0TJX8Y

Sgd 16: "peace, justice, and strong institutions". (2024). https://sdgs.un.org/goals/goal16. (Accessed: 2024-09-26)

United nations general assembly. (2024). https://www.un.org/en/ga/. (Accessed: 2024-09-26)

6 Appendix

How does the peace index (Overall Scores) vary across different continents or regions?

- A dictionary was created to map each country to its respective region (e.g., Europe, Africa, North America, South America).
- The median peace index per region was computed, and a boxplot was generated to show the distribution of peace index scores across different regions.

The following boxplots show the different regions ordered in decreasing level of peace given by the peace index score.

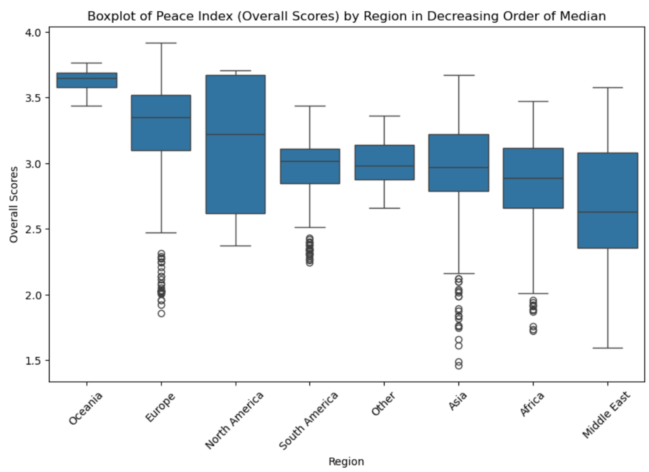

Figure 9: Boxplot of peace index by region

We can see that **Oceania, Europe and North America** lead the peace rankings with **Middle East** as the least peaceful from the rest by a comparitively large margin.