A Study of UN General Assembly Speeches and Their Relation with Security and Development

Arjuna Beuger[13402226], Sijf Schermerhorn[13179985], Patrick Spaargaren[13502573], and Kim Tigchelaar[12621803]

University of Amsterdam

Abstract. This study explores the relationship between terrorism and global development data and United Nations general debate speeches by analyzing data from the United Nations General Debate Corpus (UNGDC), the Global Terrorism Database (GTD), and the Human Development Index Database (HDID). The research aims to investigate if there is a correlation between the annual percentage of successful terrorist attempts in a country and the frequency of 'terrorism and security' related keywords in a country's annual UN speech and if a country's development level can be predicted based on its speeches during the UNCG. A negative correlation of -0.34 was found between the frequency of 'terrorism and security' keywords in UN speeches and the percentage of successful terrorist attacks, indicating that countries discussing these themes more tend to experience fewer successful terrorist attempts. For the predictive analysis, machine learning models were employed, including Multinomial Naïve Bayes, Complement Naïve Bayes, Random Forest, AdaBoost, and Support Vector Machine (SVM). The SVM performed the best, achieving an accuracy of 82.9% in predicting a country's development level based on its UN speeches. Overall, this research provides insights into the relationship between international discourse on terrorism and security and the success of terrorist attacks, as well as the predictive power of UN speeches in determining a country's development level.

Keywords: EDA, terrorism, security, development, text processing, exploratory, predictive

1 Introduction

Every September since 1946, heads of state and other high-ranking officials gather at the United Nations General Assembly (UNGA). This annual event organized by the United Nations (UN) gives its members an opportunity to present their views on international conflict and cooperation, terrorism, development and other issues concerning international politics, providing insight on government policies worldwide. The United Nations General Debate Corpus (UNGDC) is a dataset containing all speeches delivered at this event, and is updated every year. The UNGDC corpus currently consists of 7314 General Debate statements of several countries ranging from 1970 to 2023. During this assembly, various topics are discussed, with 2023's meeting focusing on themes such as financing for development, pandemic preparedness, universal health coverage, and sustainability issues [1]. Additionally, strategies to combat terrorism and other safety issues are discussed as well [2]. The semantic content of these speeches can be analyzed using text processing techniques which may be used to answer certain questions about relationships between a speech and a world event [3].
The focus of this paper will be on both terrorism and security, and global development levels. By combining the Global Terrorism Database [4] and the Human Development Index Database [5] with the UN's General Debate Corpus, this study tries to uncover relationships between these data sets to lay bare possible existing relationships between them. To narrow the study's scope, one exploratory and one predictive question was introduced:

Exploratory: *Is there a correlation between the annual percentage of terrorist attempts that succeeded and the 'terrorism and security' keyword frequency in the annual UN speeches of a country?*

Predictive: *Can the development-level of a country be predicted, based on the speeches of that country given during the United Nations General Assembly?*

We expect there to be a negative correlation between the annual percentage of terrorist attempts that succeeded and the 'terrorism and security' keyword frequency in the annual UN speeches of a country. This would mean that the more a country discusses themes regarding terrorism and security, the less successful terrorist attempts take place in that country, which would seem like a probable connection.

We also predict that the development level of a country can successfully be predicted based on the speeches of that country given during the UN's General Assembly, as it's plausible that themes discussed by countries with a low development level would defer from themes discussed by countries with a higher development level. A predictive model would pick up on those differences and be able to categorize the corresponding development level of a country correctly.

By applying statistical methods and supervised learning techniques, these hypotheses are tested and evaluated. In the remainder of the paper, we will outline our methodology, present results, and draw conclusions for both questions.

2 Methodology

2.1 Data Sources

To answer the preceding research questions, several datasets were used. The datasets were pre-processed before any statistical or analytical method has been applied. Table 1 shows an overview with how much data was used from each dataset.

United Nations General Debate Corpus The United Nations General Debate Corpus (UNGDC) [6] includes transcripts of all speeches given during the General Debate at the United Nations General Assembly (UNGA). It also includes the ISO-alpha3 code of the country that gave the speech, the year the speech occurred and its session number. From this dataset, the following 3 out of 4 columns were used: 'Year','ISO-alpha3 Code' and 'Speech'.

Global Terrorism Database The Global Terrorism Database (GTD) [4] contains information about terrorist attacks around the world, containing a variety of information about the attack. This information includes for instance the type of the attack, the name of the militant (if known) as well as the target, whether the attack was successful, the nationality of the attacker and of the target and in which country the attack took place.

During the pre-processing of this data set, the original 135 columns were reduced to a list of 20 columns (see Appendix 6 for full list). These columns were determined to be the most relevant to answer the corresponding research question.

Human Development Index Database The Human Development Index database [5], created from Human Development Reports, contains country and territory level data about several key dimensions of human development measures. It contains data about 195 countries and includes 880 columns, describing things such as human development, birth rates, carbon dioxide emissions, gender inequality, life expectancy and education. For this research, only the 'ISO3' and 'Human Development Groups' columns were used. The former to combine with the UNGDC dataset and the latter as labels for the predictive analysis.

	UNGDC	**GTD**	**HDID**
No. of used columns	3 out of 4	20 out of 135	2 out of 880
No. of rows	8480	181691	195
No. of countries	193	205	195
Time period	1970 to 2020	1970 to 2017	-

Table 1: Meta-data on datasets used

2.2 Pre-Processing the Data

Data Preparation for Exploratory Analysis The UNGDC [6] and GTD [4] were merged on country names. To use the speeches of the UNGDC for the analysis, the textual data needs to be converted to numerical values. The speeches, originally formatted as strings, were tokenized and converted into a list of individual words (excluding stop words), using the *nltk* stop words corpus and tokenizer.

To investigate the extent to which each speech focuses on security, terrorism and safety, a list with key words concerning these topics was created (see Appendix for the full list). Each instance of a key word was counted for each country per speech per year. The count of each year was summed by country, resulting in a dataframe containing the name of each country and its total sum of mentioned key words across all of their given speeches. This sum was compared to the percentage of successful terrorist attacks per country, calculated by dividing the country's total number of *successful* terrorist attacks by it's total number of attacks. Countries where the total attack count was smaller or equal to 3 were considered outliers and have been left out, because they are prone to having unusually high success rates.

Data Preparation for Predictive Analysis The UNGDC [6] and the HDI Database [5] were merged on 'ISO3' country codes, which is a universal standard to codify nations and ensures the development labels are merged to the corresponding countries correctly.

To ensure that is no "leaking" of information regarding the nation that is giving the speech, all nationality-related words have been redacted from every speech. This form of data anonymization should hurt the model's performance, but make it more robust since it won't be relying on identifying information. The data was anonymized by using the ISO codes and using the PyCountry API to get the corresponding country name, as they are more commonly called (e.g. "The Netherlands" instead of "Netherlands (Kingdom of the)". To further help redaction, we used a separate data-source containing nationalities (e.g. American or Venezuelan) [7].

The data is further processed via a pipeline that extracts features by first using a count-vectorizer, which takes a speech and turns it into a vector of word counts. These word-counts are then used by a tfidf-vectorizer to give each word in every speech an tf-idf score. The tf-idf score evaluates the importance of a word in a speech relative to the rest of the corpus. These tf-idf scores are then used as final training-features for the models.

2.3 Selected Methods

To determine whether a linear relationship between the "terrorism and security" keyword sum and the percentage of successful terrorist attacks exists and if so, in what direction, Pearson's

Correlation Coefficient was calculated.
Since there are no stakeholders involved which may require the model to be explainable, model selection was done based on performance. However, the dataset contained some class imbalance (as shown in Figure 1), which would normally result in a training bias. In order to correct for this class imbalance (where highly developed countries gave more speeches than lower developed countries), some of the selected models have been corrected for this. The RandomForestClassifier, AdaBoost and the Support Vector Machine (SVM) all contain parameters to correct for class imbalance. The Multinomial Naïve Bayes classifier does not have a parameter for correcting class imbalance. In 2003, Rennie et al. [8] proposed a solution to this problem by giving more weight to features that are more predictive of the target class, regardless of the class distribution. This algorithm, called Complement Naïve Bayes, reduces the training bias and happens to also improve performance.

2.4 Training & Experimental Setup

Out of the 8327 rows, 80 contained NaN-values. After dropping those, the 8247 rows of the remaining dataset was split into thirds, resulting in a set of 5525 rows for training and validation and a test-set consisting of 2722 rows.

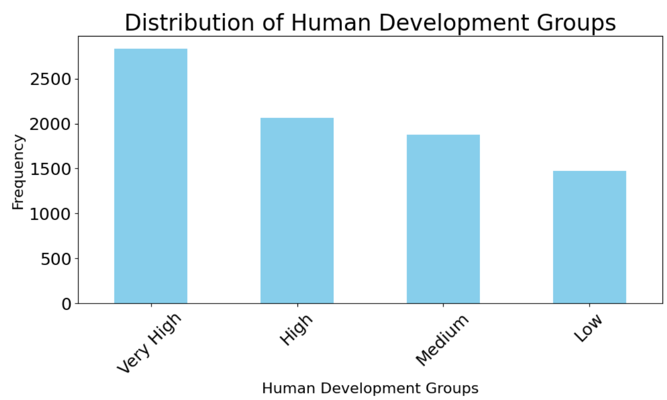

Fig. 1: Number of development group labels per class

3 Results

3.1 Exploratory Data Analysis

To visualize some relevant aspects of the GTD and UNGDC, related to *question 1*, the summary in Table 2 and the plots in Figures 2 and 3 were created.

Statistic	Number of attacks	Percentage successful
Mean	850.62	17.22
Std	2366.41	21.22
Min	0	0
25th Percentile (Q1)	17.25	1.99
Median (50th Percentile)	76	8.26
75th Percentile (Q3)	439.50	24.85
Max	21861	100.00

Table 2: Summary statistics per country

This summary shows that the average total number of attacks for each country from 1970 to 2020 is 850.62 with a standard deviation of approximately 2366.41. This standard deviation is relatively high, indicating a wide range of values. The average success percentage of all attacks is 17.22%.

Plot 2 shows a color-coded world map where the *total* number of key words used by each country (from 1970 until 2020) corresponds to a specific color (shown in the legend). The lighter the color, the higher the total frequency of key words usage of a country. Based on this plot, the countries' speeches that focus most on themes regarding terrorism and safety are located in North-America, the North of Africa and Oceania.

Plot 3 shows the progression of the yearly total number of terrorist attacks world wide. The plot shows a clear increase in the number of attacks starting from the mid-2000s, peaking around 2014 and then gradually decreasing again.

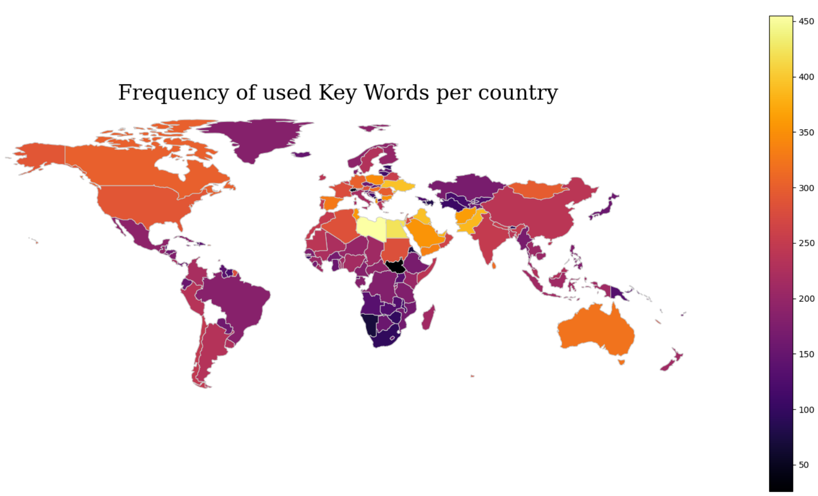

Fig. 2: Frequency of used Key Words per country

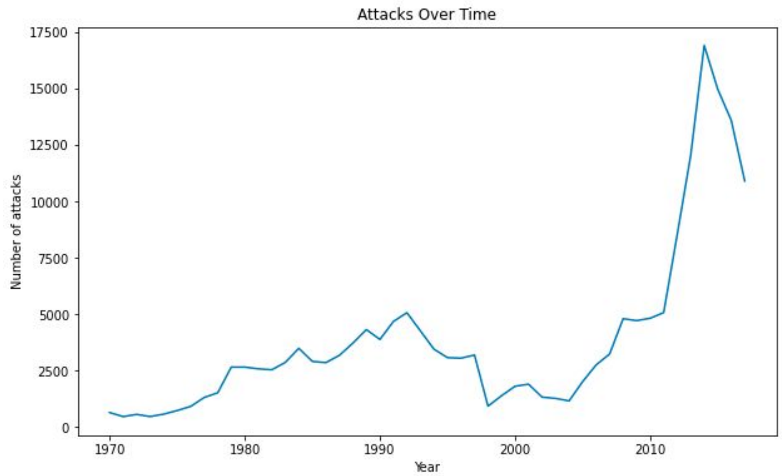

Fig. 3: Number of total terrorist attacks over time

After conducting a correlation analysis of the "terrorism and security" key word frequency of UN speeches of a country and the success percentage of terrorist attacks within this country, as explained in the Methodology section, a correlation was found. The Pearson's Correlation Coefficient for this test came out to be *-0.34*.

3.2 Predictive Analysis

The optimal hyper-parameters used in the final runs were found for all models and have been identified using grid-search [9], they have been laid out in Table 3.

Model	**Hyperparameters**
Multinomial NB	α: 0.001
Complement NB	α: 0.007
Random Forest	n_estimators: 100 max_depth: 50
AdaBoost	weak_learner: DecisionTreeClassifier n_estimators: 120 max_leaf-nodes per tree: 100
SVM	C: 10 kernel: rbf

Table 3: Hyperparameters for different models.

Inference has been run on all three models using their optimal hyper-parameters as denoted in Table 3. A Stratified 5-fold cross validation was used to ensure that the train and test sets have the same the class distribution as the original dataset [10]. For every 5-fold run, the mean and standard deviation are presented in Table 4. For each of the 5 runs, the best performing model was selected and was further evaluated on the test-set, which' results are presented in Table 5.

Model	Precision	Recall	F_1-Score	Accuracy
Multinomial NB	0.743 ±0.018	0.699 ±0.010	0.709 ±0.012	0.725 ±0.010
Complement NB	0.769 ±0.013	0.755 ±0.010	0.746 ±0.012	0.762 ±0.010
Random Forest	0.580 ±0.016	0.522 ±0.008	0.519 ±0.009	0.561 ±0.006
AdaBoost	0.654 ±0.011	0.657 ±0.010	0.655 ±0.010	0.671 ±0.011
SVM	0.804 ±0.014	0.804 ±0.011	0.804 ±0.013	0.811 ±0.012

Table 4: Validation results (mean and std.dev) of a 5-fold cross-validation

Model	Precision	Recall	F_1-Score	Accuracy
Multinomial NB	0.745	0.704	0.714	0.727
Complement NB	0.767	0.758	0.747	0.761
Random Forest	0.598	0.559	0.557	0.592
AdaBoost	0.671	0.673	0.672	0.686
SVM	0.822	0.821	0.821	0.829

Table 5: Selected models evaluation on test-set

4 Conclusion

4.1 Interpreting Exploratory Data Analysis

The correlation of *-0,34* shows that there is a slightly negative relationship between the 'terrorism and security' keyword frequency in the annual UN speeches of a country and the percentage of successful terrorist attacks in this corresponding country: The more they talk about terrorism and safety, the less successful terrorist attacks occur.

4.2 Interpreting Predictive Analysis

The results shown in Table 5, clearly suggest that the SVM performed the best and was especially consistent in its performance across the chosen evaluation measures. The Complement Naïve Bayes outperformed the Multinomial Naïve Bayes, as was expected. The Random Forest and AdaBoost models performed better than random, but still did not yield note-worthy results on this prediction task.

5 Discussion

5.1 Discussing Exploratory Data Analysis

To calculate the success percentage of terrorist attacks, countries where the total attack count was smaller or equal to 3 were left out of the calculation. Removing those countries from the dataset, reduces the risk of getting coincidental 0 or 100 percentage scores. Biased scores like this would decrease the accuracy of the model and because of this, those 'outliers' were left out.
Between the GTD and the UNGDC, there's a discrepancy between the used names for some specific countries. For example, where in one dataset the country Congo is named as "People's

Republic of the Congo", in the other data set it's called "Republic of the Congo". Additionally, some countries that appear in one dataset do not appear in the other and vice versa. This means that by merging the datasets based on country names, information about 49 countries is lost because no exact match of the country name is found between the datasets. In the future, it would be recommended that all countries not automatically matched are mapped to their correct country individually or via other means. However, the calculated Pearson's Correlation is still based on data about 145 countries and a time period of 30 years, making it still reasonably representative for the combined datasets as a whole.
The results of the calculated Pearson's Correlation demonstrate a negative correlation between the percentage of successful terrorist attacks and the amount of terrorism and safety related keywords. This corresponded with the hypothesis. It is believed that this is the case, because mentioning terrorism more means a country puts more attention towards the subject and because of this will be more likely to take measurements that will lead to prevention. However, this was an assumption, and no concrete evidence was found to support this. For this reason, it is still too early to speak of causation, and we only demonstrate a moderate amount of correlation.
The data is grouped per country and used to calculate correlations. There is a possibility that this approach can lead to oversimplification of the data, which means that interesting events or variations that occurred in different years are not considered. Because of this, the generalisation might lead to biases. However, the data was grouped per country rather than per year, because this provides a larger amount of data per country. The amount of data per instance is otherwise too small. This makes the chance of unrealistic outliers that are still used to find correlations higher.

5.2 Discussing Predictive Analysis

The SVM and the Complement NB seem capable of reliably predicting a development group label based on a processed speech. To further improve performance, the model could benefit from a larger diversity of features instead of just text.

5.3 Future work

Although there exists a (negative) correlation between the speeches at the UN and the success-rate of terrorist attacks, there may not be a causal relationship between the two. If there is, it will most likely be rather indirect. Similarly, the speeches can accurately predict the development group of a nation by its speech alone. This also does not mean that there is a causal relationship between those. In future work, any causal relationships could be investigated by performing cas

References

[1] World Economic Forum, *What is the un general assembly?* 2023. [Online]. Available: https://www.weforum.org/agenda/2023/09/un-general-assembly-what-happens/.

[2] United Nations, *United nations counter-terrorism-strategy*, 2023. [Online]. Available: https://disarmament.unoda.org/counter-terrorism/united-nations-counter-terrorism-strategy/.

[3] A. Baturo, N. Dasandi, and S. J. Mikhaylov, "Understanding state preferences with text as data: Introducing the un general debate corpus," *Research & Politics*, vol. 4, no. 2, p. 2 053 168 017 712 821, 2017.

[4] University of Maryland, *"global terrorism database"*. [Online]. Available: https://www.kaggle.com/datasets/START-UMD/gtd.

[5] S. Banerjee, *Human development world index*, Oct. 2022. [Online]. Available: https://www.kaggle.com/datasets/iamsouravbanerjee/human-development-index-dataset/.

[6] S. Jankin, A. Baturo, and N. Dasandi, *United Nations General Debate Corpus 1946-2022*, version V11, 2017. DOI: 10.7910/DVN/0TJX8Y. [Online]. Available: https://doi.org/10.7910/DVN/0TJX8Y.

[7] K. Park, *Nana dataset*, Jun. 2020. [Online]. Available: https://www.kaggle.com/datasets/bryanpark/nana-dataset/data.

[8] J. D. Rennie, L. Shih, J. Teevan, and D. R. Karger, "Tackling the poor assumptions of naive bayes text classifiers," in *Proceedings of the 20th international conference on machine learning (ICML-03)*, 2003, pp. 616–623.

[9] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, 2012, ISSN: 1532-4435.

[10] D. Berrar, "Cross-validation," in *Encyclopedia of Bioinformatics and Computational Biology*, S. Ranganathan, M. Gribskov, K. Nakai, and C. Schönbach, Eds., vol. 1, Oxford: Academic Press, 2019, pp. 542–545.

6 Appendix

Columns used from GTD

```
'Year',
'Month',
'Day',
'Country',
'state',
'Region',
'city',
'latitude',
'longitude',
'AttackType',
'Killed',
'Wounded',
'Target',
'Summary',
'Group',
'Target_type',
'Weapon_type',
'Motive',
```

```
'success',
'eventid'
```

Terrorism & Security Key Words

```
"security",
"safety",
"protection",
"defense",
"surveillance",
"law enforcement",
"public safety",
"national security",
"border security",
"cybersecurity",
"intelligence",
"emergency response",
"threat assessment",
"risk mitigation",
"resilience",
"preparedness",
"contingency planning",
"homeland security",
"safety measures",
"security protocols",
"terrorism",
"terrorist",
"counterterrorism",
"extremism",
"radicalization",
"jihadism",
"terrorist organization",
"suicide bombing",
"terror threat",
"insurgency",
"terror plot",
"counterterrorism measures",
"terror financing",
"terrorist attacks",
"radical ideology",
"violent extremism",
"terrorist recruitment",
"terrorist cells",
"counterterrorism strategy",
"counterterrorist operations"
```