# United Nations General Assembly Project

Muriël de Groot(12870331), Rinus van Grunsven(10755373),
Jonathan A. Harris(13711180), Casper Smit (11764406), Samantha Visbeek (11867175)

University of Amsterdam

**Abstract.** The United Nations General Assembly offers a unique context to study international politics, as the speeches represent its member states' official perception of and stance on current political developments. In this paper, we explore if differences in development strata, based on the Human Development Index (HDI), are reflected in the General Assembly speeches. Exploratory data analyses aimed to identify differences in word usage and frequency, speech sentiment, class-to-class reference patterns and investigate the feasibility of a multinomial classification algorithm. Our results show minimal differences in word usage between the development classes, with cosine similarities ranging from (0.93-0.99). Moreover, we find no relationship between HDI score and speech sentiment and that the proportion of references to countries in the four classes are similar across all groups. Finally, we implemented a logistic regression model with an accuracy of 71%, thus indicating some correlation between HDI class and word usage in speeches. However, the top 5 words are very general and don't relate to overarching topics or themes.

**Keywords:** United Nations · General Assembly · Natural Language Processing · Sentiment Analysis · multinomial classification.

## 1 Introduction

The United Nations (UN) is an intergovernmental organization that strives for international cooperation, human rights, environmental protection, security and maintaining peace [4]. The central policy-making organ of the UN is the United Nations General Assembly (UNGA). The General Assembly hosts a General Debate each year at which leaders of its 193 member states deliver statements on significant issues in international politics. These issues include racism, climate change, poverty, hunger, armed conflict, and other global challenges [11]. Participating countries can freely express their government's perspectives on issues they deem essential [1]. As such, the General Debate offers a unique context in which to study international politics, providing a great deal of information about the issues most salient to its member states and about their preferences [6].

In addition to the yearly UNGA, the UN is responsible for the United Nations Development Program (UNDP), a development network that promotes technical and investment cooperation. The UNDP coordinates resources to achieve sustainable development goals, including poverty reduction and HIV/AIDS. Since 1970, the UNDP has used the Human Development Index (HDI) to assess the development of countries by evaluating key metrics of human development and traditional economic growth. While the HDI metric ranges from 0 to 1, evaluated countries are categorized into four levels of development strata: Low, Medium, High, and Very High [9].

Presently, it is unknown to what extent country development, as assessed using the HDI, influences the language used or topics discussed by government leaders during the UNGA. This project aims to explore potential differences between HDI classes and the corresponding language and topics during the UN General Assembly speeches. Beyond assessing word usage and topics, speech polarity will be compared between development strata. Additionally, the project aims to elucidate patterns in how countries mentioning other countries. Lastly, the project will explore the feasibility of predicting the HDI class using a multinomial classification model.

## 2 Methodology

### 2.1 Data Sources

Three open-source data sets were used in the present study. HDI-scores of 188 countries from 1990 to 2019 were obtained from the United Nations Development Programme as a CSV file and contained the country name, year, and HDI-score [3]. The UNDP calculates HDI-score using a comprehensive method based on the geometric mean of normalized indices for each of the three dimensions:

$$HDI = (I^{Health} \cdot I^{Education} \cdot I^{Income})^{1/3} \tag{1}$$

The health metric incorporates the life expectancy at birth, education is measured by both the expected years of schooling as well the mean years of schooling, and income metric considers the Gross national income per capita.

Corresponding UN General Assembly speeches from 1990 (Session 45) to 2019 (Session 74) were collected from Harvard Dataverse in a single folder that contained the raw speeches in text files grouped by year; ISO-Alpha3 country codes were provided in the text files [1]. Lastly, a complete list of standard country or areas codes were obtained from the UN Department of Economic and Social Affairs [10]. This list contained the country name and ISO-Alpha3 codes necessary to join the HDI and UNGA data sets.

### 2.2 Data Processing

Country names were standardized using DataPrep.EDA and ISO-Alpha3 code errors were corrected in order to merge HDI and UNGA data-sets [8]. Next, HDI class labels were generated from the score based on criteria outlined by the UN: Low (HDI $\leq$ 0.550), Medium (0.550 $<$ HDI $\leq$ 0.699), High (0.700 $<$ HDI $\leq$ 0.799), and Very High (HDI $\geq$ 0.800) [9].

Before analysis, speeches were systematically cleaned and processed. First, paragraph numbers, punctuation, salutations, and references were removed. Then, standard stopwords listed in the NLTK.corpus package and custom stop words were removed [2, 7]. Lastly, the cleaned speeches were tokenized and the wordnetlemmatizer package was used to normalize the text strings [2].

### 2.3 Analyses and Multinomial Logistic Regression

Following data processing, country-to-country references within each speech were identified, and self-references were removed. Additionally, speech polarity was assessed for each speech using the Valence Aware Dictionary and Sentiment Reasoner (VADER) sentiment analysis tool [5]. Next, speeches were combined into class-specific sub-corpora, and Term Frequency-Inverse Document Frequency (TF-IDF) vectorisation was used to weigh each word based on frequency and uniqueness. The creation of word clouds per HDI class was based on TF-IDF as vectorisation removed prevalent common words from consideration.

Multinomial logistic regression was used to predict HDI class based on the speeches. Following TF-IDF vectorisation, a logistic regression model was created using the TF-IDF values of each word as predicting features and the HDI class as the target variable. The features and target values were randomly split into a training and test set with an 80:20 ratio. The model was trained on the training set, and the predictions on the test set computed the accuracy. Furthermore, the precision and recall scores for each class were computed.

We were, however interested in the features that distinguish speeches from one class to those of another class. The initial logistic regression model had an accuracy of 81%. When studying the most distinctive words for each class (i.e. the words with the highest weights), we observed that some of these words were very specific. By further investigation, it was clear that such words only appeared once in the database. The problem is that words that are only mentioned once or a few times by the same country are highly distinctive as such words will always predict the HDI score of that same country. The weights of the words do not correspond to HDI score and the form of speech.

In order to solve this, words that were only mentioned once were removed from the database. Additionally, we also removed names, as they are very exclusive to a particular country, e.g., the name of the king of Bhutan. When looking at figure 1, it is evident that the HDI scores of many countries made a significant change between 1990 and 2019. This change can lead to wrong classifications considering these country-specific names can be trained on a different HDI score, as these scores are time-dependent. Besides, such words are country-specific and do not relate HDI score to form of speech.

Therefore, the logistic regression model was trained on data-sets with and without names to compare the difference.

## 3 Results

HDI scores ranged from 0.192 to 0.957 (M=0.664, SD=0.166). The number of speeches analyzed for Low, Medium, High, and Very High HDI classes was obtained (n=1326, 1273, 1202, and 1173, respectively). The geographic distribution of HDI scores in 1990 and 2019 are presented in figure 1.

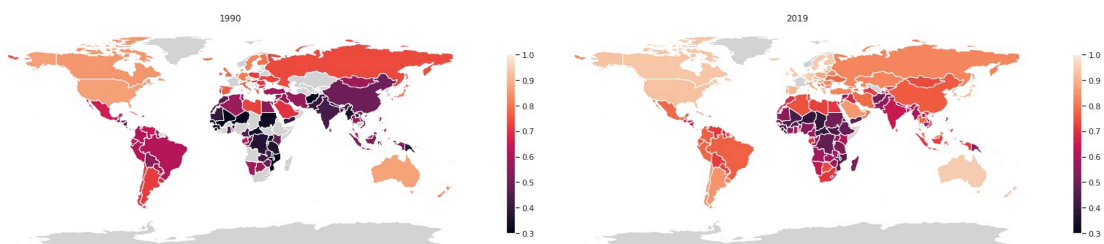


Fig. 1: The HDI score for each country in 1990 and 2019 respectively. The gray areas indicate that no data was available at that time.

Word clouds were generated per class sub-corpora of the top 30 words as defined by their vectorisation frequencies, displayed in figure 2. Less subjective than word clouds, cosine similarity was calculated to quantify likeness between the HDI sub-corpora, as can be seen in figure 3 on the left.

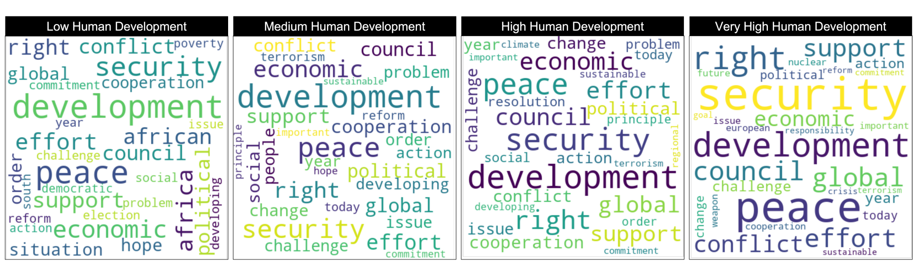


Fig. 2: Wordcloud determined by TF-IDF frequencies (n=30 words)

Sentiment analysis was performed on each speech across all HDI classes, and negative speeches were identified by their polarity score (PS < -0.05). Just 2% of countries in the Low class (n=27 speeches) gave a negative speech, followed by 4.5%, 6.4%, and 6.1% of speeches in the Medium (n=57 speeches), High (n=77 speeches), and Very High class (n=72 speeches), respectively. Polarity score means and deviations were calculated for Low (M=0.956, SD=0.274), Medium (M=0.907, SD=0.405), High (M=0.872, SD=0.464), and Very High (M=0.874, SD=0.464) classes, see figure 3 on the right.

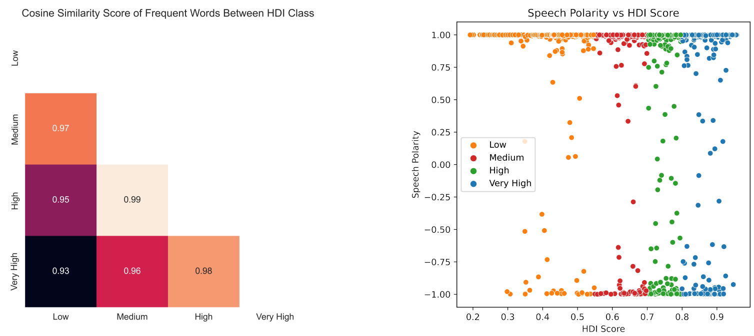


Fig. 3: Left: Cosine similarity heatmap of speeches between HDI classes. Right: Scatter plot of speech polarity as a function of HDI score

Additionally, relationships between countries referencing other countries, and their respective HDI classes, were analyzed and shown in a Sankey diagram 4. The diagram shows the relationship of the source HDI class (i.e. giving the speech) on the left side of the diagram and the target HDI class (i.e. referenced) on the right side of the diagram. Each band represents the number of speeches given by countries of one HDI class referencing countries of another class. Comparing thicknesses of the four bands emanating from a single HDI class provides a visual understanding of proportion. The Sankey diagram shows that the number of references and proportion of references to countries in the four classes is similar across all groups.

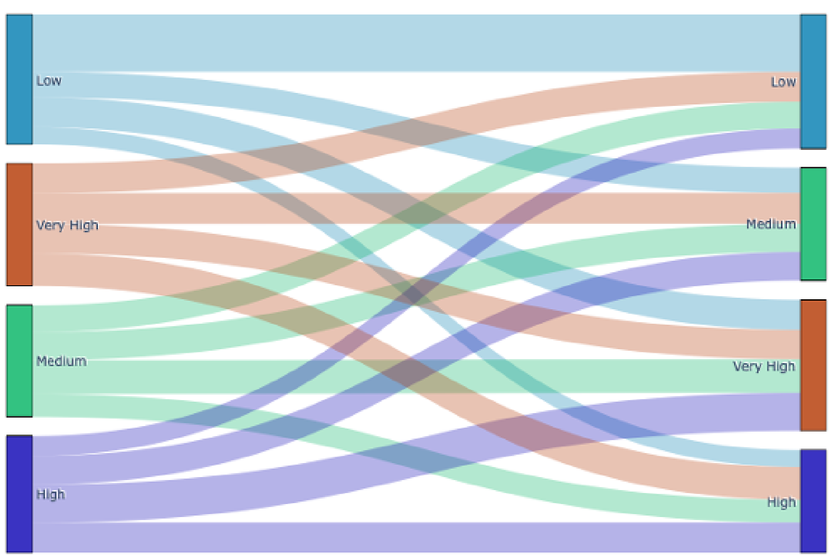


Fig. 4: Sankey diagram of countries of one HDI class (left groups) referencing countries belonging to another HDI class (right groups).

The accuracy of the logistic regression model on the database without names was 71%. The precision and recall score per class are displayed in table 2. Finally, the five words with the highest coefficient in the model are displayed in table 1.

| | | | | | |
|---|---|---|---|---|---|
| Low | election | partnering | subprime | pays | delegating |
| Medium | democrat | soccer | drudgery | graduating | penury |
| High | natal | aggressed | region | slurs | islamic |
| Very high | valued | neks | humanistic | responsibilities | rightness |

Table 1: The 5 words with the highest coefficient in the logistic regression model per class.

| | Low | Medium | High | Very high |
|---|---|---|---|---|
| Precision | 0.88 | 0.72 | 0.68 | 0.90 |
| Recall | 0.92 | 0.70 | 0.73 | 0.83 |

Table 2: The precision and recall for all hdi classes

## 4 Discussion

From the word clouds, we observe that, in general, most of the identified words are shared across HDI class sub-corpora; however, some differences are found. Both Low and Medium classes emphasise development, whereas the focus of High and Very High classes is on security.

Strong cosine similarities were observed between all classes (0.93-0.99), with the most considerable difference between Low and Very High speeches, see figure 3 on the left.

From the sentiment analysis, we observed that the percentage of negatives speeches generally increased with HDI class. Despite the minor differences in the percentage of negative speeches between classes, there does not appear to be a relationship between HDI score and Polarity Score, see figure 3 on the right.

As previously discussed, there does not seem to be a correlation between the HDI classes and the content of the speeches. Nevertheless, this was more of an exploratory analysis based on arbitrary criteria; we tried to take a more quantitative approach with the logistic regression model. The logistic model yielded an accuracy of 71%. This is a significant decrease in accuracy in comparison with the model trained on the database including names, which had an accuracy of 81%, as described in section 2.3. Regardless, an accuracy of 71% does in fact, indicate that there is some correlation between the HDI class and the topic or manner of speaking of the speeches. Ideally, investigating table 1 would reveal a different general topic for each HDI class. Here this is not the case as the top five words are very general and do not relate to a single topic. This result is most likely because the logistic regression model is not 100% accurate together with the inherent general content of the speeches.

The precision (how many predictions for this class actually belonged to this class?) and recall (how many speeches that belonged to this class were predicted as this class?) scores of the different classes produced by the model are displayed in table 2. The classes with a low and very high HDI-score scored best on both criteria, followed by the medium and high HDI-score classes. As the data was not skewed (all classes had roughly the same amount of speeches), this suggests that words that are used by countries with a low or very high HDI-score are more distinct than those of the other classes.

For future research, the logistic regression model can be improved by using bi-grams as features. Bigrams are two consecutive wore that appear in that order in a speech. Combinations with more than two words can also be used, called n-grams. Another improvement could be made by choosing a more sophisticated approach to removing all the names in the speeches. Currently, all the capitalized words are removed, which also removes e.g. words at the beginning of a sentence, illnesses and brand names. It would be interesting to note that countries with a high HDI might talk more about big corporations like Google.

## 5 Conclusion

This paper explored potential differences between HDI classes and the related topics discussed during the General Assembly speeches. Exploratory data analysis provided us with three key insights. Firstly, it can be concluded that there are no significant differences in word usage and topics between the speeches of the different development strata. Secondly, we found that the proportion of references to countries in the four classes are similar across all groups. The third critical insight of our exploratory research is that there are only minor differences in the sentiment of speeches across the HDI classes.

Besides exploratory data analysis, we also performed multi-class classification modelling. With an accuracy of 71%, it can be concluded that our logistic regression model shows some correlation between HDI class and word usage during the speeches. However, it should be noted that the top 5 words of each class are very general and do not relate to specific topics.

In total, we did not find evidence that there is no relationship between the level of development and the content of the corresponding speeches during the General Assembly.

## References


1. Alexander Baturo, Niheer Dasandi, and Slava J Mikhaylov. Understanding state preferences with text as data: Introducing the un general debate corpus. *Research & Politics*, 4(2):2053168017712821, 2017.
2. Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.", 2009.
3. UN Human Development Data Center. Human development index data, 2020.
4. Sven Bernhard Gareis. *The united nations.* Macmillan International Higher Education, 2012.
5. Clayton Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8, 2014.
6. Soo Yeon Kim and Bruce Russett. The new politics of voting alignments in the united nations general assembly. *International organization*, 50(4):629–652, 1996.
7. Anush Kocharyan. Nlp analysis of 50 years of united nations general debate speeches, 2019.
8. Jinglin Peng, Weiyuan Wu, Brandon Lockhart, Song Bian, Jing Nathan Yan, Linghao Xu, Zhixuan Chi, Jeffrey M. Rzeszotarski, and Jiannan Wang. Dataprep.eda: Task-centric exploratory data analysis for statistical modeling in python. In *Proceedings of the 2021 International Conference on Management of Data (SIGMOD '21), June 20–25, 2021, Virtual Event, China*, 2021.
9. UN. The human development report 2020 – technical notes, 2020.
10. UN. Standard country or area codes for statistical use, 2020.
11. UN. Workings of the general assembly, 2021.